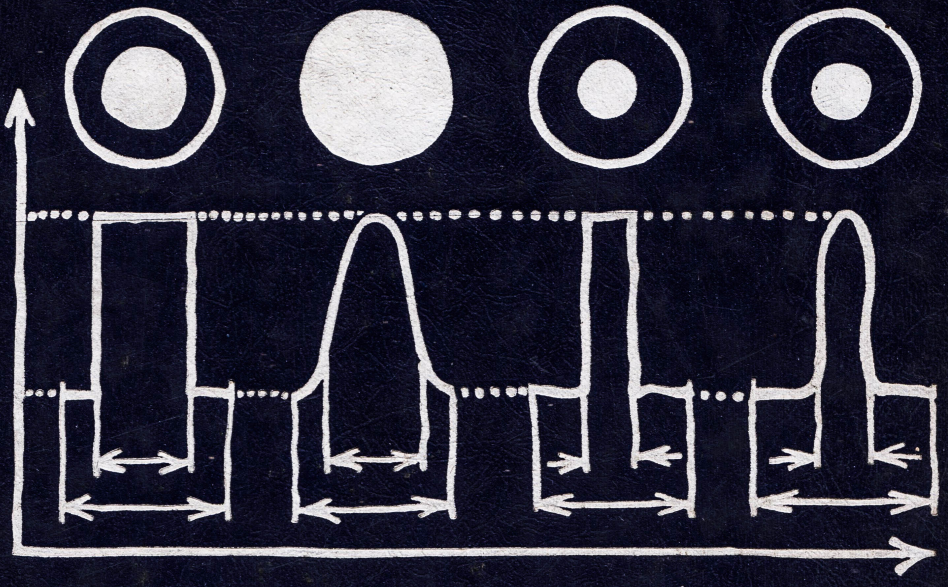
L. SHARUPICH, N. TUGOV

Opto-electronics



MIR PUBLISHERS

OPTOELECTRONICS

OPTOELECTRONICS

Л. С. Шарупич, Н. М. Тугов

ОПТОЭЛЕКТРОНИКА

Москва Энергоатомиздат

L. SHARUPICH, N. TUGOV

Optoelectronics

MIR PUBLISHERS MOSCOW

First published 1987
Revised from the 1984 Russian edition

Translated from the Russian by Peter S. Ivanov

The Greek Alphabet

A α	Alpha	I ι	Iota	P ρ	Rho
B β	Beta	K κ	Kappa	Σ σ	Sigma
Γ γ	Gamma	Λ λ	Lambda	T τ	Tau
Δ δ	Delta	M μ	Mu	Υ υ	Upsilon
E ε	Epsilon	N ν	Nu	Φ φ	Phi
Z ζ	Zeta	Ξ ξ	Xi	X χ	Chi
H η	Eta	O o	Omicron	Ψ ψ	Psi
Θ ϑ	Theta	Π π	Pi	Ω ω	Omega

На английском языке

© Энергоатомиздат, 1984
© English translation, Mir Publishers, 1987

Printed in the Union of Soviet Socialist Republics

Contents

Preface 9
Introduction 10

Chapter 1. Optical Radiation. Light-Emitting Diodes

1.1. Basic Parameters of Optical Radiation 15
1.1.1. Radiometric and Photometric Parameters 15
1.1.2. Spectral Parameters 21
1.1.3. Coherence 23
1.1.4. Polarization 28
1.2. Luminescence of Semiconductors 29
1.2.1. Kinds of Optical Radiation 29
1.2.2. Direct-Gap and Indirect-Gap Semiconductors 31
1.2.3. Injection Electroluminescence 34
1.3. Light-Emitting Diodes 38
1.3.1. External Quantum Efficiency and Radiation Loss . . . 38
1.3.2. Heterojunction LEDs 40
1.3.3. The LED as an Element of the Optron 42
1.4. Parameters and Characteristics of LEDs 43
1.4.1. Emission Characteristic 43
1.4.2. Spectral Characteristic 44
1.4.3. Optical Parameters 46
1.4.4. Electric Parameters of LEDs 51
1.4.5. Parameters and Characteristics of Industrial LEDs 53

Chapter 2. Photodetectors

2.1. Internal Photoelectric Effect 57
2.1.1. Photoconduction and Photo-emf 57
2.1.2. Photoeffect Parameters 59
2.2. Parameters and Characteristics of Photodetectors 60
2.2.1. Responsivity 60
2.2.2. Noise and Threshold Parameters 61
2.2.3. Basic Characteristics 62
2.2.4. The Photodetector as an Element of the Optron . . . 63
2.3. Photodiodes 63
2.3.1. The Principle of Action 63
2.3.2. Modes of Operation 66
2.3.3. Voltage-Current Characteristic 66
2.3.4. Energy Characteristic 68
2.3.5. Spectral Characteristic 69
2.3.6. Speed of Response 70

2.4. Photodiode Types 72
2.4.1. *PIN* Photodiodes 72
2.4.2. Schottky Barrier Photodiodes 74
2.4.3. Heterojunction Photodiodes 76
2.4.4. Avalanche Photodiodes 77
2.5. Current-Amplifying Photodetectors 79
2.5.1. Phototransistor 79
2.5.2. Composite Phototransistor 80
2.5.3. Photothyristor 81
2.5.4. Comparison of Characteristics 82
2.6. Photoresistors 82
2.6.1. Photoconduction 83
2.6.2. Parameters and Characteristics 84
2.7. Photodetectors Using the Charge-Coupled Concept 85
2.7.1. MOS Capacitor 86
2.7.2. Charge Packet and CCD Gate Operation Modes . . . 88
2.7.3. Data Recording and Readout 90

Chapter 3. Optrons

3.1. Block Diagram of an Optron 93
3.1.1. Optron Properties 93
3.1.2. Block Diagram 94
3.1.3. Light Source 95
3.1.4. Optical Medium 96
3.1.5. Photodetector 98
3.1.6. Optron Classification 98
3.2. PCP Parameters 100
3.2.1. Input and Output Parameters 100
3.2.2. Transfer Parameters 100
3.2.3. Isolation Parameters 103
3.3. Basic Types of Optrons 104
3.3.1. Diode Photocoupled Pairs (D PCPs) 104
3.3.2. Transistor Photocoupled Pairs (T PCPs) 106
3.3.3. Thyristor Photocoupled Pairs (Y PCPs) 107
3.3.4. Resistor Photocoupled Pairs (R PCPs) 107
3.3.5. Optoelectronic Integrated Circuits (OICs) 107
3.3.6. Special Optrons 108
3.3.7. Parameters and Characteristics of Industrial Photon-Coupled Pairs 109
3.4. Application of Optrons 114
3.4.1. Classification of Optrons by the Fields of Application 114
3.4.2. Application of Optrons in Digital and Pulse Devices 114
3.4.3. Analog Applications of Optrons 119
3.4.4. Optron Switches 123
3.4.5. Application of Special Optrons 123
3.5. Reliability of Optoelectronic Devices 125
3.5.1. Reliability Parameters 125
3.5.2. Physical Causes of Failure 126
3.5.3. Reliability Criteria 127
3.5.4. Radiation Stability 128

Chapter 4. Coherent Optoelectronics

4.1. Physical Processes in Lasers 129
4.1.1. General 129
4.1.2. Light Amplification 130
4.1.3. Population and Population Inversion 132

4.1.4. Lasing 133
4.1.5. Lasing Threshold 134
4.2. Laser Structure, Parameters, and Modes of Operation . . . 135
4.2.1. Laser Structure 135
4.2.2. Basic Parameters of Lasers 137
4.2.3. Modes of Operation 137
4.3. Semiconductor Lasers 138
4.3.1. Laser Amplification in Semiconductors 138
4.3.2. Injection Lasers 140
4.3.3. Heterojunction Lasers 142
4.3.4. Electron-Beam Lasers 143
4.4. Laser Types 144
4.4.1. Gas Lasers 144
4.4.2. Solid-State Lasers 146
4.4.3. Comparison of Lasers 147
4.5. Laser Modulators 148
4.5.1. Laser Modulation 148
4.5.2. Physical Principles of Laser Modulation 149
4.5.3. Optical Modulators 151
4.5.4. Thin-Film and Semiconductor Modulators 153
4.5.5. Deflectors 155
4.5.6. Space-Time Modulators (STMs) 156
4.6. Fiber-Optic Communication Lines 157
4.6.1. FOCL Block Diagram 158
4.6.2. Fiber Optic Guide 159
4.7. Holographic Data Systems 161
4.7.1. Principle of Holography 161
4.7.2. Holographic Memory 162
4.7.3. Holographic Recording and Readout Devices 163

Chapter 5. Fabrication of Optoelectronic Devices

5.1. Fabrication Processes 164
5.1.1. Semiconductor Wafer Production 164
5.1.2. Epitaxy 165
5.1.3. Oxidation 167
5.1.4. Doping 168
5.1.5. Photolithography 170
5.1.6. Assembling 172
5.2. Fabrication of LEDs 174
5.2.1. Production of Single Crystal Semiconductor Compounds 174
5.2.2. Fabrication of LED *pn* Junctions by Diffusion . . . 178
5.2.3. *PN* Junction Formation by Vapor Phase Epitaxy (VPE) 182
5.2.4. *PN* Junction Fabrication by Liquid Phase Epitaxy (LPE) 185
5.2.5. Ohmic Junction Fabrication 188
5.2.6. Typical Processes of LED Fabrication 191
5.3. Fabrication of Photodetectors 196
5.3.1. Materials 196
5.3.2. Fabrication of *pin* Photodiodes 196
5.3.3. Fabrication of Schottky-Barrier Photodiodes 198
5.3.4. Fabrication of Light-Emitting Heterodes 199
5.3.5. Fabrication of Phototransistors 200
5.3.6. Fabrication of Photothyristors 202
5.3.7. Fabrication of Photoresistors 203
5.3.8. Fabrication of Avalanche Photodiodes 204
5.4. Fabrication of Photon-Coupled Pairs 205

5.4.1. The Elements of a Photon-Coupled Pair (PCP) . . . 205
5.4.2. PCP Assembling and Packaging 207
5.5. Fabrication of Fiber Guides 208
5.5.1. Fabrication Techniques 208
5.5.2. Fiber Guide Structures 210
5.6. Fabrication of Semiconductor Lasers 212
5.6.1. Laser Structure 212
5.6.2. Techniques of Laser Fabrication 213
Index 218

Preface

Optoelectronics is the branch of science and technology which deals with the study of photo-electronic and electro-optic phenomena in solids and the utilization of solid-state devices based on these phenomena. Modern optoelectronics benefits from the achievements in a number of related fields, first of all semiconductor and quantum electronics. Optoelectronics has sprung from electronics. Now it both supplements the latter and gradually grows into a separate domain of science and technology.

The building blocks of optoelectronics are various radiators, photodetectors, and coupling elements. At present the output growth rate of these devices greatly exceeds that of many other classes of semiconductor devices.

This book aims to introduce the reader to the elements of most important physical phenomena used in semiconductor optoelectronics and also the physical properties, parameters, and characteristics of basic optoelectronic, primarily semiconductor, devices.

In describing each class of devices, the text covers in succession the physical processes, principles of action and structures of the devices, their parameters, characteristics, fields of application, and basic types. Last chapters consider basic manufacturing methods and production cycles in the technology of fabrication of optoelectronic devices.

Introduction

Modern electronics essentially relies on electronic phenomena in solids, primarily, in semiconductors, and utilizes semiconductor devices based on these phenomena. Electronic devices are noted for high sensitivity to control signals, functional flexibility, high speed, and versatility. There seemed to be no technical tasks which electronics would have been unable to cope with in the future. But developments in electronics have revealed serious limits on the practical application of electronic phenomena and, hence, electronic devices. In particular, electronic systems have low noise immunity because of the strong mutual interaction between elements at the attained level of the packing density of integrated circuits. Thus, today an airborne computer weighs 20 to 30 kg, whereas the mass of cables connecting the computer with sensors and indicators reaches a huge value of above 5 tons. What causes an increase in the mass of cables is the need to protect data transfer channels from noise and the adverse effect of channels on one another.

Noise gives designers of control and automatic systems much trouble as it calls for exceptionally reliable shielding of electronic equipment and trunk lines used for transferring various signals.

Electronic transfer channels also present other serious drawbacks. They primarily spring from the growing shortage of copper and large expenses on the production of cables, auxiliary equipment, and the maintenance of underground cable ducts. Municipal authorities of many large cities voice concern of the alarming situation that exists now—the available ducts get congested, but the layout of new ones in the network of the now existing complex communication lines involves insuperable difficulties and calls for huge expenses.

Particular difficulties of electric isolation arise where electronic systems have an appreciable difference in electric power from element to element. Thus a typical electronic device—a transducer—consists of a control section based on integrated microcircuits (ICs) and a power section that directly operates into a load and switches over the power of large values. In evidence is the need to match the IC-based control section having a low noise immunity with the

power section responsible for high levels of noise. So, electronics today faces acute problems of both electric isolation and increase in the information capacity of channels.

Modern electronics has not solved and cannot ultimately solve the problem of the total microminiaturization of electronic equipment. Such traditional components as transformers, relays, and variable resistors are poorly compatible with integrated circuits because they are large in size and have moving parts. The above components incorporated into electronic devices account for a major share of the volume and mass of the devices, power consumed, number of failures, and the cost that sometimes reaches 90 to 95% of the total cost of devices. Microminiaturization is particularly urgent for devices of data input and output. Thus, the most widely used component of this type—a cathode ray tube—requires a high voltage (over 1 kV) for its operation, has a large mass and appreciable dimensions.

Modern microelectronics has practically reached the upper theoretical limit as regards both the speed of data transfer and the memory capacity of computers. At the same time the body of information to be handled grows continuously. The present memories of computers must be comparable in storage size to that of the man's brain (about 10^{13} bits), and the speed of data transfer must be equal to 10^{10}-10^{12} bits per second for computers and 10^{8}-10^{10} bits per second for communication lines. Such values of memory capacities and speeds are practically impossible to achieve using devices based only on electronic phenomena.

Both research and practice have shown that the problems stated above are solvable through the joint utilization of electronic and optical phenomena in semiconductors, namely, by way of conversion from electronics to optoelectronics. Optoelectronics deals with the study of electronic, optical, photo-electronic, and electro-optic phenomena and the use of devices and systems based on these phenomena.

An *optoelectronic device* depends for its operation on electromagnetic radiation in the optical range. On the whole, optoelectronic devices are rather diverse and fall into the following basic groups.

1. Light radiators such as lasers and light-emitting diodes (LEDs).
2. Photoelectric detectors. These are photoresistors (photoconductors) and *pn* junction photodetectors.
3. Radiation control devices, such as modulators, light beam deflectors, and controllable transparencies.
4. Data display devices such as indicators and display panels.
5. Optrons, whose varieties are known as photon-coupled pairs (PCPs), source-detector pairs, optoisolators, etc. used for electric isolation between the circuit components.
6. Optical communication channels and optical memories (storages). These are fiber-optic light conduits and storage media (both permanent and reversible).

The basic components of any optoelectronic system are light radiators which are the sources of coherent radiation (lasers) and incoherent radiation (mainly LEDs). The devices of coherent (laser) optoelectronics sharply differ from the devices of incoherent optoelectronics in the principles of generation, transmission, and detection of signals. The devices of these two groups also differ in the requirements placed on their parameters; they perform different functions and reflect different levels of technical development. So, we can differentiate between two types of optoelectronics: (1) coherent optoelectronics concerned with coherent emission of radiation and (2) incoherent optoelectronics, whose devices rely on incoherent emission of radiation.

Coherent optoelectronics offers the following principal advantages that determine its development.

1. High information capacity of an optical channel because optical frequencies extend to the range of 10^{13} to 10^{15} GHz and are 10^3 to 10^5 times higher than radio frequencies harnessed for data transmission. A small wavelength of optical radiation ensures a high data recording density of optical memory devices (in the order of 10^8 bits per cm^2).

2. The possibility of dual modulation (both temporal and spatial) of an optical signal since it embodies such properties as frequency, phase, polarization, and amplitude. This, in turn, offers a means of handling complete images and changing optical signals in space.

3. The possibility of further microminiaturization of functional components and conversion to media with distributed parameters by virtue of the fact that the wavelengths of optical radiations are short. Thus, there is an ideal basis for the creation of devices of functional electronics.

The stage of creating the building blocks of coherent optoelectronics will probably take no less than 10-15 years.

The level of development of incoherent optoelectronics is far higher: most types of photodetectors, LEDs, and optrons are now commercially available and industry has recently started to turn out optoelectronic ICs in quantities. At present, incoherent optoelectronics acts as a supplement of the highly developed electronics and microelectronics. It substantially extends the functional capabilities of electronic equipment.

The circuits and systems of incoherent optoelectronics offer the following principal advantages.

1. Almost ideal electric isolation of circuit elements due to optical coupling provided by photons which do not possess electric charges. At present, the basic elements used in electronics to provide electric isolation between units and circuits are transformers and capacitors, whose shortcomings are known fairly well: design and technological incompatibility with integrated circuits; presence of

feedback paths through parasitic reactances; narrow frequency range; large mass and size. Optical coupling almost completely obviates these shortcomings and can ensure electric isolation between circuit sections differing in potentials by more than 1 000 V. The isolation resistance of optical coupling can reach 10^{16} ohms and the transfer capacitance can be as small as 10^{-4} pF.

2. Unidirectional transfer of a signal from radiator to photodetector. The reverse transfer of a signal from the output to the input, i.e. parasitic feedback between photodetector and radiator, is practically nonexistent. This feature simplifies the structure of the optronic system and, besides, facilitates the design procedure. The design engineer can separately analyze individual stages, calculate their parameters, and then use the results to design a multistage system.

3. The possibility of obtaining information in the form of light signals used to keep control over the status of individual units.

4. A low level of noise since optical and electrical signals pass through two separate channels, optical and electrical.

Optical isolation offers a number of advantages which are lacking in purely electronic circuits. Thus, it becomes possible to control high voltages up to hundreds of kilovolts by low voltages since it is only the dielectric strength of insulation that determines the upper voltage limit. The same also concerns electric power. Optical isolation enables easy coupling of circuits operating at different frequencies through conversion, for example, direct current into alternating current. Electrically separated circuits sharply increase the switching capabilities in complex systems owing to the absence of interaction between optical channels. Circuits feature enhanced noise immunity since optical channels are not subject to electromagnetic noise.

Optoelectronic devices can operate from a dc source, and their upper frequency limits ranging into 10^7-10^8 Hz totally depend on the speed of a radiator and a photodetector.

It should be pointed out that optoelectronics makes a tremendous impact on many branches of industry and brings about both quantitative and qualitative changes in electronics, radio engineering, communications engineering, computer engineering, automatics, and cybernetics. Optoelectronics marks a qualitatively new stage in the development of electronic engineering.

Let us briefly look into the history of optoelectronics and enumerate the basic stages of its development.

1. In 1864 James Clerk Maxwell formulated his theory of electromagnetic waves and laid the foundation of the wave theory. He revealed the electromagnetic nature of radio waves and optical vibrations.

2. In 1873 W. Smith discovered the photoconductive effect.

3. In 1917 A. Einstein set up the photoelectric equations for induced radiation.

4. In 1940 V.A. Fabrikant described the principle of amplification of radiation on its passing through a medium.

5. In 1947 Dennis Gabor invented holography—the method of obtaining the 3-d image of an object.

6. In 1954 A.M. Prokhorov, N.G. Basov (USSR), and C.H. Townes (USA) created the first prototypes of gas lasers.

7. In 1956 injection electroluminescence was discovered.

8. In 1960s gas, solid-state, and semiconductor lasers appeared.

9. In 1970s various light-emitting diodes and other optoelectronic devices became commercially available.

Optoelectronics is thus a very young branch of science and technology which, as an independent domain, came into being not over two decades ago.

Modern optoelectronics gave birth to integrated optics (microoptics). A direct impetus to the development of this new branch was the search for the creation of optical communication systems. These systems use discrete components such as lasers, detectors, lenses, mirrors, and other devices externally connected to one another to enable data processing in electrical channels. The aim pursued by integrated optics is to do away with double conversion of a signal (optical-to-electrical signal conversion, and vice versa) and use only an optical signal as the carrier of information. Integrated optics seeks the ways similar to microelectronic manufacturing methods for the fabrication of composite elements of microoptic devices on the common substrate in the single production cycle.

Chapter 1

Optical Radiation. Light-Emitting Diodes

1.1. Basic Parameters of Optical Radiation

1.1.1. Radiometric and photometric parameters. The principal difference between optoelectronic and purely electronic devices is that the former use or produce the energy of optical radiation apart from the electric energy. By its nature, optical radiation, just like radio wave radiation, is the emission of energy in the form of electromagnetic waves, whose spectrum appears in Fig. 1.1.

Optical radiation covers the range of wavelengths λ extending from 1 nm to 1 mm. The optical range includes the infrared (IR) region of wavelengths from 0.78 μm to 1 mm, visible region (0.38 to 0.78 μm), and ultraviolet (UV) region (1 nm to 0.38 μm).

The basic quantity of optical radiation is the energy it transfers. The parameters, or quantities, associated with the transfer of radiant energy by optical radiation are known as *radiometric* (*radiant*) parameters. To estimate the visible radiation in the optical range, i.e. the flux that produces a visual sensation, one has to pass to *photometric* (*luminous*) parameters, or quantities.

The human eye as a receptor of visible radiation (light) responds differently to the light of different wavelengths. Its response is the highest at a wavelength of 0.555 μm, but sharply decays with distance from this maximum point. At the boundaries of the visible region (0.38 μm and 0.78 μm) the eye sensitivity practically drops to zero. Fig. 1.2 illustrates the spectral luminous efficiency of the eye for photopic vision (curve *1*) and scotopic vision (curve *2*). The function V_λ adopted by the Commission Internationale de l'Eclairage (CIE) and known as the CIE spectral luminous efficiency function is the result of averaged data obtained from many experiments on human observers.

The radiometric and photometric quantities are in essence identical, but describe radiation in different wave bands and have different units of measure. Quantitative characteristics of visible light define the flux in terms of a visual sensation; in the IR and UV regions of optical radiation unperceived by the eye, the parameters characterize

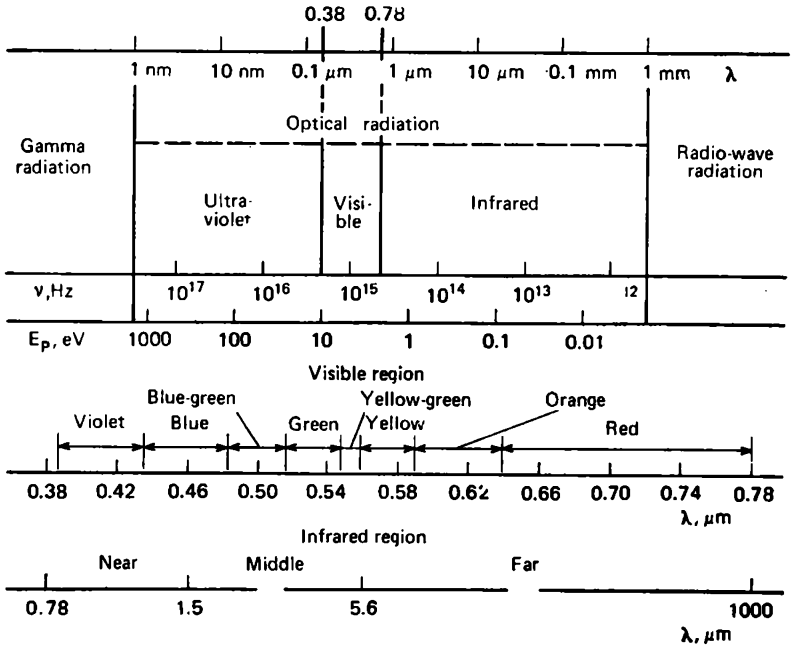


Fig. 1.1. Electromagnetic wave spectrum

the energy transferred by radiation. Radiant and luminous parameters certainly exhibit a proportionate relationship.

Before turning to the parameters of radiation, consider the geometrical notion of a solid angle which is a measure of opening of a certain cone (Fig. 1.3).

The *solid angle* Ω is the ratio of the area σ of the sphere's segment to the square of the radius r of the sphere, $\Omega = \sigma/r^2$. The unit of a solid angle is the steradian (sr). This is the solid angle, whose vertex lies at the sphere's center, which encloses on the sphere an area equal to the square of the sphere's radius, $\Omega = 1$ sr if $\sigma = r^2$. The total solid angle is obviously equal to 4π sr.

There are five basic radiometric quantities used to describe the sources and detectors of radiant energy: the radiant flux (or power) Φ_e, radiant intensity I_e, radiant exitance M_e, radiance L_e and irradiance, or dose-rate, E_e. Correspondingly, five photometric (luminous) quantities are used to describe visible radiation: luminous flux Φ_v, luminous intensity I_v, luminous exitance M_v, luminance L_v, and illuminance E_v. By definition, the radiometric and photometric parameters are analogous, but units of measurement of these quantities are different. Consider the radiometric and photometric parameters in more detail.

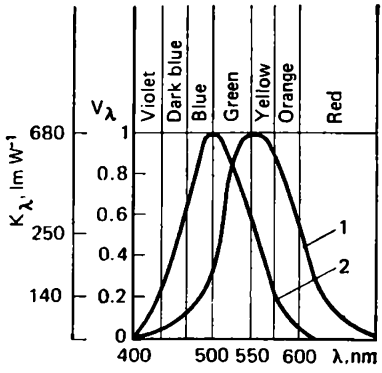


Fig. 1.2. CIE spectral luminous efficiency curves for photopic (*1*) and scotopic (*2*) vision

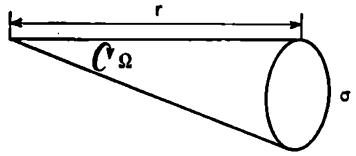


Fig. 1.3. Defining the solid angle

Radiant flux and luminous flux. Let a point radiation source, whose dimensions are much smaller than the distance from this source to the point of observation, lie at the vertex of a solid angle (Fig. 1.4*a*). This source emits electromagnetic waves in all possible directions. The radiant flux Φ_e is the mean flow of radiant energy Q_e carried by the flux of photons in a unit time, i.e. the radiant power expressed in watts and transferred by the flux within a specified solid angle Ω.

The radiant flux Φ_e falling on to the human eye causes the eye to perceive it only selectively, i.e. the eye only responds to the luminous flux Φ_v (see Fig. 1.2). The relation between the liminous flux expressed in lumens (lm) and radiant flux expressed in watts takes the form

$$\Phi_v = K_\lambda \Phi_e = K_{max} V_\lambda \Phi_e \qquad (1.1)$$

where K_λ is the luminous efficacy (luminosity), lm W^{-1}; K_{max} is the maximum luminous efficacy; and $V_\lambda = K_\lambda / K_{max}$ is the luminous efficiency.

Formula (1.1) has naturally the meaning only for the visible region of optical radiation; beyond this region, $K = 0$ and $\Phi_v = 0$. For the normal photopic vision at a wavelength of 555 nm corresponding to the maximum of the eye sensitivity, the radiant flux of 1 W is equivalent to the luminous flux of 680 lm; i.e. the luminous efficacy K (555 nm) $= K_{max} = 680$ lm W^{-1}. The values of K_λ are also given in Fig. 1.2. For modern semiconductor radiators—light-emitting diodes—generating light of different colors, the values of K_{max} are the following: 60, 135, 540, and 640 lm W^{-1} for 655-nm red, 635-nm red, 585-nm yellow, and 565-nm green LEDs respectively.

Thus, the higher the coefficient K_{max}, the greater is the share of

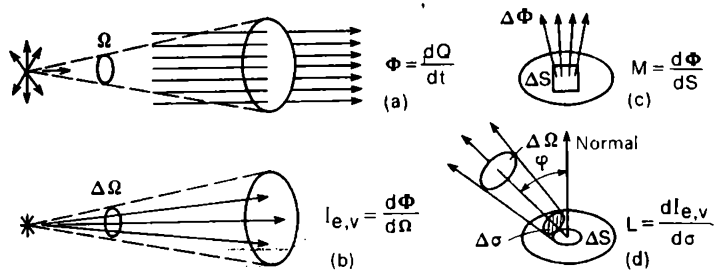


Fig. 1.4. Defining radiometric and photometric parameters

the source radiation that the eye can perceive and the more efficient is the given radiator.

The *radiant intensity* I_e or luminous intensity I_v of a source of radiation in the given direction is the ratio of the radiant flux or luminous flux, $\Delta\Phi$, to a solid angle $\Delta\Omega$:

$$I_{e,v} = \Delta\Phi/\Delta\Omega \tag{1.2}$$

To obtain a more accurate value of $I_{e,v}$, we proceed to the limit:

$$I_{e,v} = \lim_{\Delta\Omega\to 0} \Delta\Phi/\Delta\Omega = d\Phi/d\Omega \tag{1.2a}$$

If the source emits uniformly, then

$$I_{e,v} = \Phi/4\pi$$

where Φ is the total radiant flux.

Thus the radiant (luminous) intensity specifies the spatial flux density (Fig. 1.4*b*).

The unit of radiant intensity is W sr^{-1} and that of luminous intensity is the candela (cd). In the International System (SI) of units, the candela is the basic unit of measurement specified with the aid of a standard source of radiation (1 cd = 1 lm sr^{-1}).

Exitance and radiance. If a source of radiation differs from the point one, the quantities exitance (emittance) and radiance are used to specify the radiation emanating from a unit area of the emitting surface (Fig. 1.4*c*).

The radiant exitance M_e or luminous exitance M_v is equal to the ratio of the flux $\Delta\Phi$ emitted from the area ΔS in all directions within a solid angle of 2π sr to this area:

$$M_{e,v} = \Delta\Phi/\Delta S \tag{1.3}$$

The point radiation density (exitance at a point) can be found by proceeding to the limit:

$$M_{e,v} = \lim_{\Delta S\to 0} \Delta\Phi/\Delta S = d\Phi/dS \tag{1.3a}$$

The unit of radiant exitance is the watt per meter squared (W m^{-2}). The unit of luminous exitance is the lumen per meter squared (lm m^{-2}). This is the emittance of the surface of 1 m^2 in area that uniformly emits the luminous flux of 1 lm.

The radiance L_e or luminance L_v in a given direction is the ratio of the radiant (luminous) intensity dI within a unit solid angle cutting out the elemental area of the radiant surface, dS, to the area $d\sigma$ of projection of this elemental area on to the plane normal to the direction of radiation (Fig. 1.4*d*):

$$L_{e,v} = dI/d\sigma = dI/dS \cos \varphi \tag{1.4}$$

The unit of radiance is the watt per steradian-meter squared (W sr^{-1} m^{-2}). The unit of measure of luminance is the candela per square meter (cd m^{-2}). This is the luminance of the surface emanating the flux of 1 lumen from the area of 1 m^2 in the direction normal to this surface.

The *irradiance* (*dose-rate*) E_e or *illuminance* E_v of a surface is the ratio of the radiant (luminous) flux $\Delta\Phi$ impinging on an element of this surface to the area of this element, ΔS,

$$E_{e,v} = \Delta\Phi/\Delta S \tag{1.5}$$

To specify irradiance (illuminance) at a point, we should proceed to the limit:

$$E_{e,v} = \lim_{\Delta S \to 0} \Delta\Phi/\Delta S = d\Phi/dS \tag{1.5a}$$

Thus, irradiance is the flux per unit area of a radiation detector. The unit of irradiance is the watt per meter squared (W m^{-2}) and the unit of illuminance is the lux (lx); 1 lx = 1 lm m^{-2}. Irradiance (illuminance) has the same unit of measure as radiant (luminous) exitance. But these quantities differ widely in physical essence.

In a number of cases it is required to convert luminous parameters to radiant parameters, and vice versa, to estimate the performance of radiators. Thus the certificates on light-emitting diodes commonly specify the emission in terms of the luminous intensity or luminance. But these parameters prove impracticable for LEDs used in, say, the devices of optical communication and signalling. We can pass to radiant quantities by employing, as mentioned earlier, the luminous efficacy which, in the general case, is equal to

$$K = \frac{\Phi_v\,(\text{lm})}{\Phi_e\,(\text{W})} = \frac{I_v\,(\text{cd})}{I_e\,(\text{W sr}^{-1})} = \frac{L_v\,(\text{cd m}^{-2})}{L_e\,(\text{W sr}^{-1}\,\text{m}^{-2})}$$

For example, given the luminance intensity, $I_v = 3.5$ mcd and luminous efficacy, $K = 150$ lm W^{-1}, we obtain the radiant intensity, considering that 1 cd = 1 lm sr^{-1}:

$$I_e = \frac{3.5 \times 10^{-3}\ \text{cd}}{150\ \text{lm W}^{-1}} = 23.5\ \mu\text{W sr}^{-1}$$

But if a LED used in the signalling system serves as a visual display, we can convert the specified radiant parameters to luminous parameters if the value of K is nown.

The parameters of optical radiation, their short definitions, calculation formulas, SI units and symbols appear in Table 1.1.

Table 1.1. Basic Radiometric and Photometric Quantities

Radiometric quantity		Definition	Formula	Photometric quantity	
name and symbol	SI unit			name and symbol	SI unit
Radiant flux Φ_e	W	Rate of transfer of radiant energy	$\Phi = \frac{dQ}{dt}$	Luminous flux Φ_v	lm
Radiant intensity I_e	W sr^{-1}	Flux within unit solid angle	$I = \frac{d\Phi}{d\Omega}$	Luminous intensity I_v	cd = lm sr^{-1}
Radiant exitance M_e	W m^{-2}	Flux per unit area of emitting surface	$M = \frac{d\Phi}{dS}$	Luminous exitance M_v	lm m^{-2}
Radiance L_e	W sr^{-1} m^{-2}	Flux within unit solid angle, related to the area of emitting surface and directed at an angle to the normal on to the plane of this surface	$L = \frac{dI}{dS \cos \varphi}$	Luminance L_v	cd m^{-2} = lm sr^{-1} m^{-2}
Irradiance E_e	W m^{-2}	Flux per unit area of the photodetector surface	$E = \frac{d\Phi}{dS}$	Illuminance E_v	lx = lm m^{-2}

Radiation in the optical range can be defined by both wave and corpuscular entities and parameters. It can be expressed in terms of the flux of photons. The energy of a photon is given by the relation

$$E_P = h\nu = hc/n\lambda \tag{1.6}$$

where h is Planck's constant equal to 4.13×10^{-15} eV s; c is the velocity of light in a vacuum, equal to 3×10^8 m s^{-1}; n is the refractive index of an optical medium; ν and λ are the frequency (Hz) and wavelength (m) of optical radiation respectively.

Let N photons pass through a unit area S in a unit time. The expression for Φ then takes the form

$$\Phi = NE_P S = N \frac{hc}{n\lambda} S \tag{1.7}$$

The shorter the emission wavelength, the stronger the flux at the specified number N of photons. It is useful to express the number of photons through radiometric parameters:

$$N = \frac{\Phi}{SE_P} = 5 \times 10^{16} \lambda \frac{\Phi}{S} \quad (1.8)$$

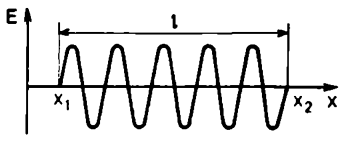


Fig. 1.5. Wave train radiated by a single atom

Here N is expressed in $1/m^2s$, λ in μm, and Φ/S in W m^{-2}.

Consider now some properties of optical radiation, which determine the areas of application of optoelectronic devices.

1.1.2. Spectral parameters. One of the basic parameters of a radiator is its *bandwidth*, i.e. the frequency band or the wave band within which the given device emits radiation. The notion of *monochromaticity* is used to estimate the bandwidth of a device.

The mechanism of emission of optical waves will be given detailed treatment below. Here we shall only point out that the emitting particles are excited atoms; an excited atom having an excess of energy passes to a lower energy state and gives up this energy as it emits an electromagnetic wave. The process of transition from the higher to the lower energy state lasts for a time Δt of about 10^{-8} s and the atom takes as much time to emit the wave. The atom thus emits not a continuous wave but a wave packet, or wave train (Fig. 1.5). The length of such a wave packet is $l = x_2 - x_1 = c_0 \Delta t \approx 3$ m, while the wavelength of an optical vibration is about 10^{-6} m, so millions of optical waves can be present along the length of the wave packet.

Figure 1.5 illustrates the oscillations of only the electric component E of an electromagnetic wave. The magnetic component is commonly omitted from the graphs of optical vibrations. This is because the magnetic component exerts a substantially weaker effect on free electrons than the electric component. But it should be remembered that both of these components are the part and parcel of the electromagnetic wave. Under no conditions can we produce a wave that would contain only one component of the field.

In the general case, optical radiation propagating from incoherent sources is a nonsinusoidal wave which can be represented as a series of harmonic (sinusoidal) components. The amplitudes and initial phases of various harmonic components determine the spectrum of optical radiation.

Optical vibrations can be represented by both periodic and aperiodic functions. Let us now take a look at the properties of periodic and aperiodic functions and consider the methods of resolving these functions into harmonic components.

The function $f(t)$ is a periodic function of time if the following

condition is met:

$$f(t) = f(t + nT)$$

where T is the period of the function; and n is an integer, either positive or negative.

The periodic function displays the property of having one and the same value at any two moments of time separated by the time interval equal to the period T. The periodicity of the phenomenon described by the function $f(t)$ must exist infinitely long, from $t = -\infty$ to $t = +\infty$.

The periodic function can be expressed as a harmonic progression or a Fourier series

$$f(t) = A_0 + \Sigma_{k=1}^{\infty} A_k \cos(2\pi k\nu_0 t + \varphi_k) \qquad (1.9)$$

where A_0 is the mean of the function over the period or the constant component; $A_k \cos(2\pi k\nu_0 t + \varphi_k)$ are harmonic (sine wave) oscillations with amplitudes A_k, initial phases φ_k, and frequencies $k\nu_0$.

The sinusoid of frequency $\nu = k/T$ at $k = 1$ is called the first or fundamental harmonic; at $k = 2, 3, \ldots$, we respectively have the second, the third harmonic, etc. A sinusoidal oscillation lasts infinitely long and has time-invariable parameters such as an amplitude A_k, frequency $k\nu_0$, and initial phase φ_k. A nonsinusoidal, but periodic, oscillation can be represented as a sum of sinusoidal components in the form of a Fourier series. The greater the number of these components, the more accurate is the expansion of the function $f(t)$ into a Fourier series. The graphical illustration of the periodic function resolved into harmonic components is given in Fig. 1.6*a*.

In optoelectronics the expansion into a Fourier series applies for the study of the spectrum of optical radiation: the radiation is represented as a sum of sinusoidal components (harmonics) which are the function of the coordinates of a point on the surface of a radiator. Under the optical spectrum described by the periodic function one understands the spectrum of harmonics which form a complex function. By the level of radiation is meant the value of any quantity X_e proportional to the square of an amplitude of electromagnetic oscillation. Depending on what parameter is used to define the spectrum, one distinguishes between the levels of a radiant flux, radiant intensity, irradiance, etc. The spectrum is graphically presented as a plot of the levels of radiation laid off on the y-axis against frequencies laid off on the x-axis. The spectrum of a periodic function can be linear, with the lines spaced at the same interval (Fig. 1.6*b*).

The spectrum of an aperiodic function is described by the spectral concentration (density) of a certain radiant quantity X_e, which is the ratio of dX_e in a frequency interval $d\nu$ to this interval. The spectrum of aperiodic oscillations contains all frequencies and displays a continuous dependence of the spectral concentration on frequency. This spectrum is called continuous (Fig. 1.6*c*).

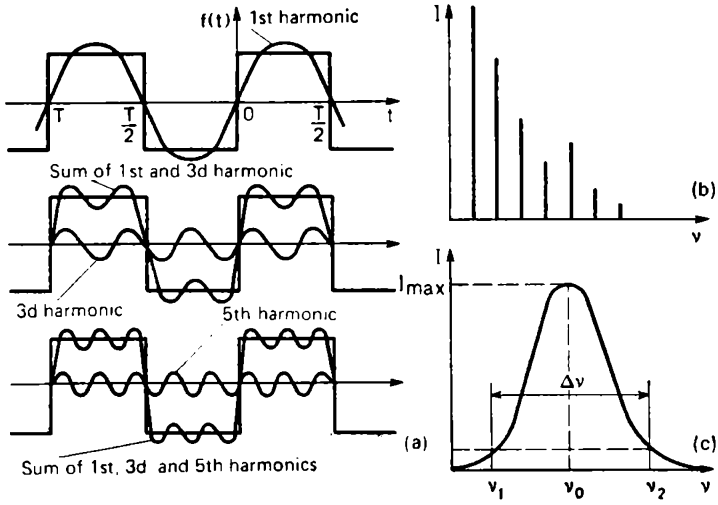


Fig. 1.6. Spectral analysis of optical radiation
(*a*) expansion of function into harmonic components; (*b*) and (*c*) types of radiation spectrum

The spectrum of radiation (emission bandwidth) is defined by the degree of monochromaticity, $\mu = \Delta\nu/\nu_0$, where $\Delta\nu$ is the bandwidth, and ν_0 is the central frequency (see Fig. 1.6*c*).

When $\Delta\nu/\nu_0 = 0$, the radiation is ideally monochromatic; when $\Delta\nu/\nu_0 \ll 1$, it is quasi-monochromatic. The parameter μ thus characterizes the deviation of monochromatic radiation from the ideal pattern whose bandwidth is equal to zero. From the viewpoint of quantum mechanics, for ideally monochromatic radiation there can correspond the radiation of energy resulting from the transition of electrons between infinitely thin energy bands. Ideal monochromaticity does not exist in nature. In reality, radiation occupies a certain interval in the optical range. A fairly narrow spectral interval can be defined by one frequency or wavelength. When radiation spreads over an extended region of the spectrum, the degree of monochromaticity is expressed as the log of the ratio between the cutoff frequencies or wavelengths of the spectrum:

$$\mu = \ln \frac{\nu_2}{\nu_1} = 2.3 \log \frac{\nu_2}{\nu_1} = 2.3 \log \frac{\lambda_1}{\lambda_2} \qquad (1.10)$$

where λ_1 and λ_2 are the cutoff wavelengths of the spectral range of optical radiation; ν_1 and ν_2 are the cutoff frequencies of the spectral range. In the formula, $\lambda_2 < \lambda_1$ and $\nu_2 > \nu_1$.

1.1.3. Coherence. This is an important property of optical radiation from the viewpoint of application of an optical signal as a carrier of information. It is coherence that determines one of the

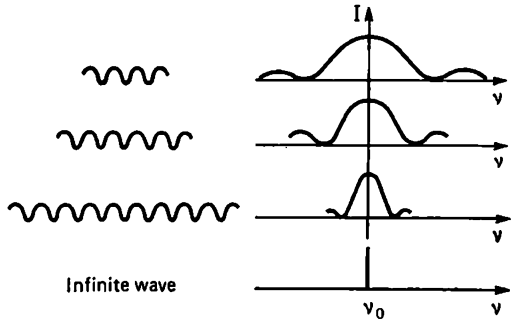


Fig. 1.7. Relation between wavelengths and radiation spectrum

basic trends in optoelectronics (laser optoelectronics). *Spatial* (*transverse*) *coherence* specifies the agreement between the phases of oscillation at various points in space at the same moment of time. *Temporal* (*longitudinal*) *coherence* is the measure of the constancy of phase in one and the same point in space at different moments of time. A change in the phase of a wave can only be observed when this wave interferes with another wave coherent with the first. The higher the coherence of a radiation source, the sharper and brighter will be the interference pattern of two coherent waves.

An absolutely coherent source is a monochromatic point radiator. But neither monochromatic nor point radiators exist in nature. The degree of deviation from monochromaticity is defined by the temporal coherence function; the deviation of the dimensions of a source from those of a point source is defined by the spatial coherence function.

As mentioned above, atoms emit energy in the form of wave packets. The length of the emitted wave packet and the spectrum of radiation are related to each other. The longer the wave packet, the narrower the spectrum of radiation (Fig. 1.7). Only in the limiting case, when the wave is infinitely long, can we obtain a monochromatic radiation of frequency ν_0.

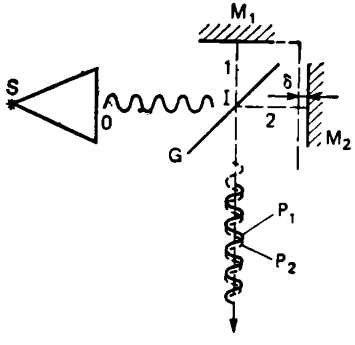


Fig. 1.8. Defining the temporal coherence

Consider the interference of polychromatic radiation in a simple interferometer consisting of a semitransparent glass G and two plane mirrors M_1 and M_2 (Fig. 1.8). Place the mirrors normal to each other so that each forms an angle of 45° with the plate G.

The polychromatic radiation from a source S is a wave packet extending from the objective lens O and plate G. At the point I the plate G separates the wave packet into two packets, one of which travels a path *1*, and the other a path *2*. If the mirror M_2 lies in the plane shown by a dash line in Fig. 1.8 (a distance δ shifted from its original position), the packet P advancing along the path *2* will cover a shorter distance in the forward and the backward direction than the packet moving along the path *1*. At the exit from the interferometer the wave packet P_1 will lag behind the packet P_2 in time by τ, called the lag time that characterizes the phase relationship between the two waves in time. If the lag time is much shorter than the duration of each of the two wave packets, then at the exit from the interferometer these two packets will practically coincide with each other and produce a very sharp and well defined fringe pattern. In this case the temporal coherence is said to exist between the vibrations.

Consider now a sufficiently extended radiation source consisting of a large number of point sources such as atoms of the emitting material. If the source is small in size, all the atoms give the same interference pattern, and the common pattern looks similar to that produced by an individual atom but is much brighter. This is the case of spatial coherence. With an increase in the diameter of the source, the distances between the interference bands produced by various atoms will be different due to the wave path-length difference, i.e. phase mismatch between the waves in space. The maxima and minima of interference patterns do not coincide with one another. The bands become more smeared and the pattern less sharp. This is the case when the source is said to generate a partially coherent light in space. A further increase in the diameter of the source results in a spatially incoherent emission. So, the spatial coherence of radiation varies with the size of a radiation source.

The phase relationship, i.e. the coherence of two radiation sources, is defined by the mutual coherence function $\gamma_{12}(\tau)$. When $\gamma_{12}(\tau) = 0$, the radiant intensity at a certain point is equal to the total intensity of oscillations propagating from each source, and these oscillations are incoherent. If the oscillations are coherent, the resultant intensity can be higher or lower than the total intensity (equal to zero in a particular case). The mutual coherence function is given by

$$\gamma_{12}(\tau) = \frac{I_{max} - I_{min}}{I_{max} + I_{min}} \tag{1.11}$$

where I_{max} and I_{min} are the maximum and the minimum intensity in the interference pattern.

The temporal coherence function specifies mutual coherence at the given point in space. This function, expressed in terms of the spectral

concentration of $\Phi_{e,\,v}$, has the form

$$\gamma(\tau) = \left| \frac{\int_0^\infty \Phi_{e,\,v} \exp(-2\pi v\tau)\, dv}{\int_0^\infty \Phi_{e,\,v}\, dv} \right| \tag{1.12}$$

The mutual coherence function having the lag time τ equal to zero is known as the spatial coherence function $\gamma_{12}(0)$. The higher the degree of spatial coherence, the smaller the beam divergence (angle of spread). Incoherent radiation is omnidirectional and has a wide angular spread.

A radiation source comprises a host of atoms, and the radiation at the given moment from the given point is a certain sum of waves generated by various atoms. The phases of waves and the relationship between these phases must be treated here as random functions of time. Such random processes and the degree of their correlation are mathematically described by the theory of correlation. Let us briefly look at some statements of this theory.

A random function is the function that in the course of time can assume one specific form or another unpredicted beforehand. The implemented form of the function that results is known as the realization of the random function. The repeated experiments can give a set of these forms (values) of the random function.

Consider what values a random function can take at a definite moment of time. We shall deal here with a certain entity of random values, called a *cross section of the random function*. The parameters of the random function in any of its cross sections are the probabilistic characteristics such as mathematical expectation, or variance.

The *expectation* of a random function $y(t)$ is a function $m_y(t)$ which is equal to the variance of the corresponding cross section of the random function at each value of the argument t. The expectation m_y of a random function is the weighted mean value of this function, i.e. the mean value calculated with consideration for the 'weight' of each value that is proportional to the probability of this value:

$$M[y(t)] = m_y(t) = \sum_{i=1}^{n} y_i P_i$$

where y_i is a possible value of the random variable or the random function in its cross section; and P_i is the probability of this value.

The expectation $m_y(t)$ of the random function $y(t)$ is a certain mean function which gives a measure of the spread of the random function about its mean (Fig. 1.9).

The expectation describes a random function (random process) in any cross section, but it does not reveal the relationship between

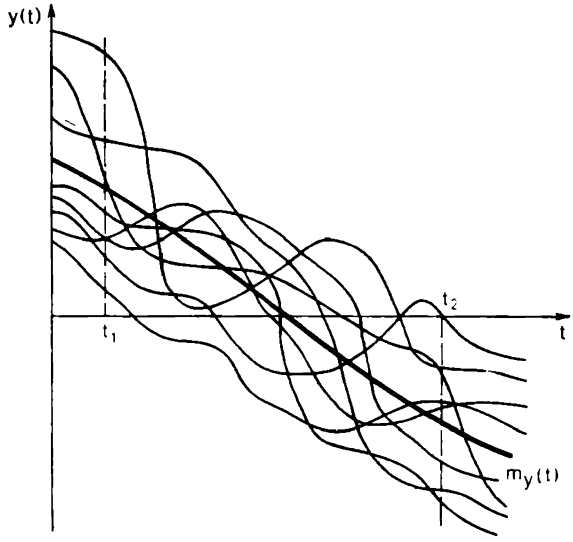


Fig. 1.9. Mathematical expectation of a random function

the cross sections at different moments of time. This relationship, and interrelation in general, of random processes can be defined by a correlation function and cross-correlation function.

The correlation function of a random variable $y\ (t)$ is a non-random function of two arguments, $K_y\ (t_1,\ t_2)$, which, given the values of each pair of arguments t_1 and t_2, is equal to

$$K_y(t_1,\ t_2) = M[y'(t_1)\, y'(t_2)] \tag{1.13}$$

where $y'\ (t_1) = y\ (t_1) - m_y\ (t_1)$, and $y'\ (t_2) = y\ (t_2) - m_y\ (t_2)$. When $t_1 = t_2$, the correlation function goes into the variance of a random function:

$$K_y = D(t) = D[y(t)]$$

The degree of correlation, r_{12}, is the correlation function presented in normalized form:

$$r_{12}(t_1,\ t_2) = \frac{K_y(t_1,\ t_2)}{\sqrt{D_y(t_1)\, D_y(t_2)}}$$

The degree of correlation is indicative of the extent of correlation: where $r_{12} = 0$, correlation is absent; when $r_{12} = 1$, correlation practically converts into a functional relationship between variables: corresponding to each value of one of the variables there will be a definite value of another variable.

It should be pointed out in particular that the mutual coherence function coincides with the envelope of the normalized correlation

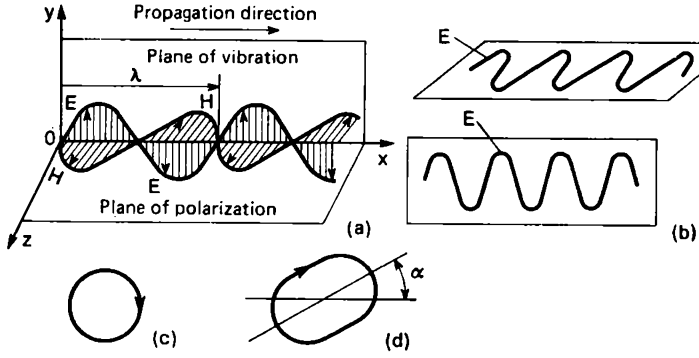


Fig. 1.10. Polarization of optical radiation
(*a*) plane of polarization and plane of vibration; (*b*) longitudinal and transverse polarization; (*c*) projection of circular polarization; (*d*) projection of elliptically polarized radiation

function (degree of correlation) for the electric component of an electromagnetic wave.

1.1.4. Polarization. This is an important radiation property that finds a number of uses in optoelectronics. Polarized radiation is the radiation that has a certain preferential direction of vibration of an electric vector. Given the preferential direction of vibration, we can speak of the degree of polarization. It is only an ideally monochromatic wave that can have ideal polarization, i.e. one, strictly definite, direction of vibration of an electric vector.

Electromagnetic radiation displays both an electric and a magnetic vector, one normal to the other. In the case of plane-polarized radiation, the magnetic vector lies in the plane of polarization. The electric vector lies in the plane of vibration which is normal to the plane of polarization (Fig. 1.10*a*). Some possible directions of vibration are also shown in the figure. The resultant curves determine the direction of polarization—these are the lines connecting the ends of a large number of electric vectors extending from various points on the x axis.

Figure 1.10*b* illustrates two examples of monochromatic radiation propagating in the x direction. These are the examples of linearly polarized radiation. It is possible to visualize an infinite set of directions of linear polarization, each forming an angle α with the horizontal axis z. This angle is called the azimuth of polarization.

A more complex example of polarization is the elliptical polarization (in particular, the circular one) defined by the azimuth, direction of rotation, and ellipticity (Fig. 1.10*c* and *d*). Any other form of polarization can be regarded as a particular case of elliptical polarization.

A basic parameter of optical radiation is the radiating (emissive)

power, i.e. the energy the radiation transfers in a unit time. If a radiator emits energy in the form of pulses, the notions of pulse power and mean pulse power are used to estimate radiation. The pulse power P_p is the total energy Q_p in a pulse related to the pulse length t_p:

$$P_p = Q_p/t_p \tag{1.14}$$

The mean pulse power is given as

$$P_{p\,m} = P_p\,(t_p/T) \tag{1.15}$$

where T is the pulse repetition period.

The power related to a unit spectral interval over which radiation occurs is known as the spectral concentration (density) of the power.

In conclusion, let us compare incoherent and coherent radiators. Incoherent radiators emit in a wide spectral band where it is generally possible to separate an interval of monochromatic radiation whose energy is, however, very low. An incoherent radiator is an omnidirectional source which can give a directional radiation only at the expense of some loss in the emitted energy. A coherent radiator emits radiation in a narrow spectral band, which propagates within a small solid angle.

1.2. Luminescence of Semiconductors

1.2.1. Kinds of optical radiation. There are two kinds of optical radiation — thermal radiation caused by heating and luminescent emission resulting from luminescence of some materials. The spectral radiant exitance of various bodies heated to a certain temperature is defined by the Planck formula known as the Planck law:

$$M(\lambda, T) = \frac{2\pi hc_0^2}{\lambda^5}\left[\exp\left(\frac{hc_0}{kT\lambda}\right) - 1\right]^{-1} \tag{1.16}$$

where k is the Boltzmann constant; λ is the emission wavelength; h is the Planck constant; T is temperature; and c_0 is the vacuum velocity of light (in the formula, we omit the subscript e at M for simplicity[1]).

The curves at $T = 6\,000$ and 300 K describe emission from the Sun and the man respectively (Fig. 1.11). At temperatures above 2 500 K, a portion of the spectrum of thermal radiation falls within the visible region. The dash line in Fig. 1.11 connects the maxima of thermal radiation, for which the following values hold: $\lambda_{max} \approx \approx 3\,000/T$ and $M(\lambda_{max}, T) \approx 1.3 \times 10^{-11} T^5$ (here the wavelength λ is expressed in μm, temperature T in kelvins K, and M in W cm^{-2}).

Incandescent lamps use a tungsten filament enclosed in a sealed glass bulb and heated by electric current to 2100-2300 K to generate

[1] Translator's note.

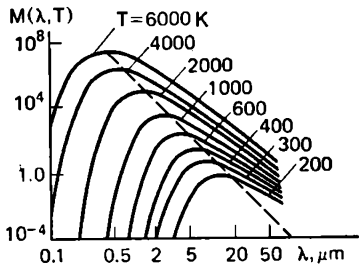


Fig. 1.11. Spectral characteristics of hot bodies

thermal radiation. These radiators emit totally incoherent light in a very wide continuous spectrum, offer a low luminous efficiency, and have long thermal lag and short service life. Besides, they are poorly stable to mechanical influences, large in size, and incompatible with integrated circuit technology. This explains why heat radiators have received limited applications in optoelectronics.

Modern radiators (emitters) use luminescent materials to produce luminescent emission of radiation.

Luminescence as a physical phenomenon has been known for more than 50 years. An example of natural luminescence is the luminescence of fireflies, rotten wet wood, etc. By modern conception, luminescence is the electromagnetic nonthermal radiation whose duration appreciably exceeds the period of optical vibrations. Thus, in distinction from emission of glowing bodies, luminescent emission does not require heating, though the energy of one form or another is of course necessary to induce luminescence. Besides, unlike the phenomenon of light scattering, luminescence persists for some time after disconnection of the source of excitation energy [1]. In other words, a luminescent material retains for some time the absorbed energy before its conversion partially to light and partially to heat.

Depending on the form of energy used to excite luminescence, one distinguishes between photoluminescence, electroluminescence, and other types. Solids, liquids, and gases are able to luminesce. Optoelectronics largely uses impurity wide-bandgap semiconductors as luminescent materials.

Consider the simplest mechanism of luminescence—photoluminescence which arises under radiation (Fig. 1.12). A quantum of exciting radiation falling on the semiconductor can be absorbed by a luminescent center (transition *1*) or by a crystal lattice of the host material (transition *2*). The semiconductor is said to store light-generated carriers, i.e. the energy in the form of separated charges which can give a certain number of photons during recombination that occurs at the points of crystal defects of the translation lattice, such as interstice stranger atoms and vacant lattice sites, called recombination centers. In luminescent semiconductors the recombination centers can be radiative and quenching. Radiative centers are able to con-

[1] Luminescence that lasts for a time in excess of 10^{-8} s after absorption of energy is known as phosphorescence. If the luminescence time is shorter than 10^{-8} s, the process is known as fluorescence.

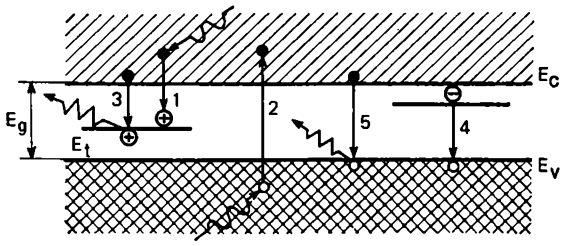


Fig. 1.12. Photoluminescence in semiconductors

vert the energy evolved during recombination into light (transitions *3* and *5*). At the quenching centers the recombination energy is commonly converted into heat (transition *4*), i.e. this energy does not contribute to luminescence. The process of luminescence thus consists of two basic stages. At the first stage the excitation energy generates carriers, so the semiconductor stores the 'light energy'. It is this stage—the stage of separation of carriers—that determines the type of luminescence. At the second stage—the stage of recombination—the semiconductor gives up the stored energy. This stage is common to all types of luminescence.

It is thus clear from the photoluminescent process described above that for the semiconductor to be able to store a certain amount of 'light energy', the semiconductor atoms should be brought out of thermodynamic equilibrium, i.e. they must be driven to the excited state under irradiation.

At the second stage when the semiconductor releases the stored energy, the emission of quanta takes place as an electron passes to a lower energy level either directly or via a recombination trap.

1.2.2. Direct-gap and indirect-gap semiconductors. At the first stage of luminescence electrons can make various transitions: band-to-band (interband), band-to-impurity, and impurity-to-impurity transitions. Band-to-band transitions can be direct and indirect. These two types of transition stem from the dependence of the energy of an electron on its momentum (see Figs. 1.13 and 2.2). The electron momentum P_e is equal to the product of the electron mass m_e and speed v:

$$P_e = m_e v \tag{1.17}$$

The direct transition occurs without a change in the electron momentum. The indirect transition involves a change in the electron momentum that is made up for by the momentum of an emitted or absorbed photon.

By the law of conservation of momentum, during emission or absorption of a photon the following equality must hold:

$$P_{e1} = P_{e2} \pm k_P \tag{1.18}$$

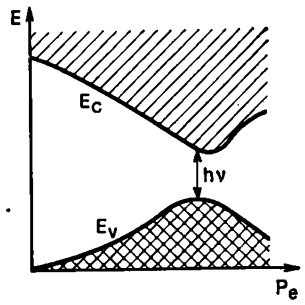


Fig. 1.13. Electron energy versus momentum for direct electron transitions

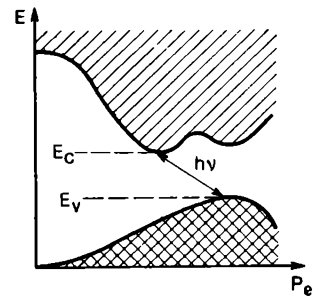


Fig. 1.14. Electron energy versus momentum for indirect electron transitions

where P_{e1} and P_{e2} are the initial and finite momenta of the electron, respectively; and k_P is the photon momentum.

Since the photon velocity is equal to the velocity of light, then $k_P = m_P c_0$, where m_P is the photon mass related to the wavelength by the de Broglie equation

$$m_P = h/\lambda c \tag{1.19}$$

The photon momentum is then equal to

$$k_P = m_P c = h\nu/c \approx E_g/c \tag{1.20}$$

where E_g is the bandgap (gap energy).

When $E_g \approx 1$ eV, we have $k_P \ll P_{e2}$, i.e. the electron momentum may be considered invariable in the direct transition ($P_{e1} \approx P_{e2}$), which corresponds to the transition along the vertical between the peak of the valence band, E_v, and the valley of the conduction band, E_c (see Fig. 1.13).

Electrons can also make indirect transitions from the valence to the conduction band. These transitions involve changes in the electron momentum ($P_{e1} \neq P_{e2}$). Apart from the photon and electron, a third particle must then participate in the process of energy absorption to carry away a portion of the momentum (Fig. 1.14). For indirect transitions the law of conservation of momentum has the form

$$P_{e1} = P_{e2} \pm k_P \pm k \tag{1.21}$$

where k is the momentum of a third particle, for example, a phonon.

The basic materials of semiconductor radiators are GaAs and its ternary compounds GaAlAs and GaAsP. These are direct-gap semiconductors, i.e. the semiconductors with the allowed optical transitions directly from band to band. Each act of recombination of a charge carrier during the direct transition results in the emission of

a photon whose wavelength (μm) is written in the form

$$\lambda \approx 1.23/E_g \qquad (1.22)$$

Thus, in order to fulfil the law of conservation of momentum in direct transitions (which must be met, as must be the law of conservation of energy, to enable any electronic transition), there is no need for a third particle to participate in recombination, apart from the electron and photon. For this reason the probability of direct optical transitions is high, and direct-gap semiconductors are efficient luminescent materials.

In indirect-gap semiconductors, such as gallium phosphide GaP, the conduction band valley proves shifted along the axis of momentum. The radiative recombination of an electron with a hole occurs only via a certain complex which takes up an excessive momentum and thus a fraction of energy. Nevertheless, radiative recombination can effectively proceed via appropriate impurity centers in two steps: first, a carrier of one sign resides in an impurity center and then this carrier recombines with a free carrier of the opposite sign. In GaP, for example, donor-acceptor complexes (Zn^+-O^-) or neutral traps (an N atom instead of a P atom in the GaP lattice) can act as the centers of radiative recombination.

It should be noted that self-absorption of radiation is much stronger in direct-gap semiconductors than in indirect-gap counterparts.

Table 1.2 lists the materials used for the manufacture of semicon-

Table 1.2. Materials of Semiconductor Radiators

Type	Semiconductor	Gap energy, eV	Interband transition
$A^{III}B^{V}$	GaAs	1.45	Direct
	GaP	2.25	Indirect
	AlAs	2.16	Indirect
	GaN	3.5	Direct
$A^{II}B^{VI}$	ZnSe	2.7	Direct
	ZnS	3.8	Direct
	ZnO	3.2	Direct
	ZnTe	2.3	Direct
	CdS	2.5	Direct
	CdSe	1.8	Direct
	CdTe	1.6	Direct

Note. Along with binary compounds, solid solutions, largely ternary compounds such as GaAlAs, GaAsP, and InGaP are very popular. The structure of formulas for ternary compounds shows what atoms of the elements substitute one another in the crystal lattice. The values of energy gaps and the structure of energy bands in solid solutions depend on the percentage of components forming a solution.

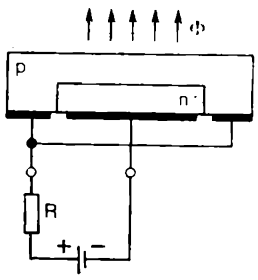


Fig. 1.15. Electroluminescence from a *pn* junction

ductor radiators and also gives the gap energy E_g for each material. As is evident from the table, direct-gap semiconductors enjoy more uses. The choice of the value of bandgap E_g depends on the operating emission wavelength λ_{em} of a radiator in the optical range.

From expression (1.22) we have

$$E_g \approx 1.23/\lambda_{em} \qquad (1.22a)$$

Consequently, for radiators to be able to operate in the visible range (0.38 to 0.78 μm), the semiconductors should have an energy gap of 1.5 to 3.0 eV. This requirement rules out at a stroke the application of germanium and silicon, whose technology is well developed, and motivates the need for conversion to $A^{III}B^{V}$ compounds, their solid solutions, and other appropriate materials.

Semiconductor devices usually rely on electroluminescent effects to generate optical radiation. The energy of excitation of electroluminescence is taken up from an electric field. There are two types of electroluminescence:

(1) injection electroluminescence which appears in the forward-biased *pn* junction (Fig. 1.15);

(2) avalanche electroluminescence which develops in materials exposed to strong fields, close in strength to the fields causing electric breakdown of the *pn* junction. The first type has found most extensive uses in radiators.

1.2.3. Injection electroluminescence, i.e. generation of optical radiation in the *pn* junction, combines two processes: injection of carriers and luminescence proper. The injection process is called upon to generate nonequilibrium charge carriers and thus to store 'light energy'. The mechanisms of injection are given detailed treatment in the books on the physics of semiconductors and semiconductor devices. Here we shall dwell on some of the features of injection in emitting structures.

Under equilibrium conditions the Fermi levels of two homogeneous *p*- and *n*-type semiconductors brought in contact must be coincident. This causes the bending of energy bands and the formation of a potential barrier, as is clear from the band diagram illustrated in Fig. 1.16. To estimate the motion of positively charged holes and negatively charged electrons in the *pn* junction field, let us apply the known illustrative analogy: visualize the holes which drift from right to left in the direction of the field E as light bubbles floating up to the top of the valence band and also picture the electrons which drift from left to right in the direction opposite to that of E

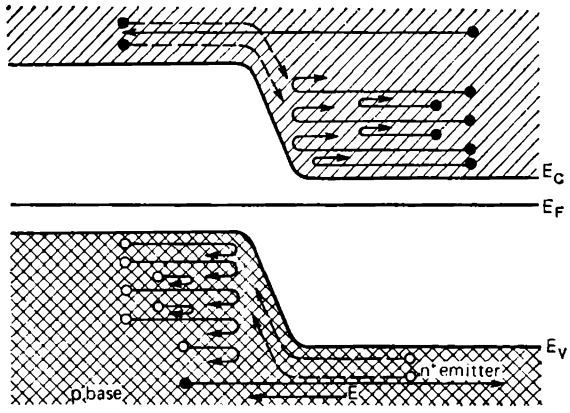


Fig. 1.16. Motion of carriers across a *pn* junction

as heavy balls rolling down the slanting bottom of the conduction band. Most of the holes move rightward from the *p* layer rich in these carriers to the junction region, but cannot overcome the potential barrier and return to the *p* layer as they penetrate into the junction for a certain depth. The holes in the *n* layer readily 'float' up from the bottom of the valence band regardless of the energy of the *p* layer and form the drift flow moving from right to left. This flow is balanced out by the counter flow of holes diffusing from the *p* layer and having high energy enough to overcome the potential barrier. The pattern of electron motion is similar: electrons freely roll off the *p* layer into the *n* layer, forming the drift flow of electrons. This electron flow is cancelled by the flow of higher-energy electrons moving from the *n* layer. The forward voltage applied to the junction lowers the potential barrier and thus gives rise to the diffusion currents of both holes and electrons, i.e. increases the injection of minority carriers—holes into the *n* layer and electrons into the *p* layer.

The *pn* junction commonly emits light only from its one region (*p* region in Fig. 1.15). It is obviously desirable that the number of injected carriers should be maximum precisely in the emitting (active) *p* region. For this purpose the amount of donor impurity introduced in the *n* region is made higher than that of acceptor impurity in the *p* region. The weakly doped *p* region, called the base, has a high resistance. The strongly doped *n* region, denoted by n^+ and called the emitter, has a low resistance. So, injection of carriers in the emitting structure practically proceeds in one direction, from the emitter into the base which emits light.

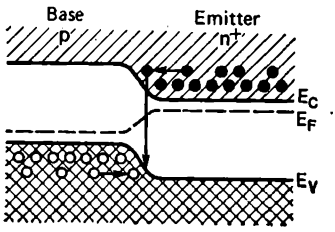


Fig. 1.17. Recombination of carriers across a *pn* junction

As mentioned earlier, the materials of light-emitting structures must have a wide forbidden band. An appreciable, and even predominant, current that appears in these structures is the one caused by the recombination processes in the space charge region of the *pn* junction (Fig. 1.17). The larger the bandgap, the higher is the potential barrier and the more intensive is the recombination of electrons in the *pn* junction. This recombination usually occurs at deep luminescent centers and results in the generation of a phonon *hq* (generation via recombination centers *2*, as illustrated in Fig. 1.18). These electrons are thus lost in nonradiative recombination, and the recombination current I_{rec} they produce decreases the injection efficiency of 'radiative' electrons.

The useful component of current providing the radiative recombination in the *p* base is the electron current I_n injected by the emitter. The efficiency of injection depends on the extent to which the current I_n differs in magnitude from the total current I. The measure of injection efficiency is the injection ratio

$$\gamma = \frac{I_n}{I} = \frac{I_n}{I_n + I_p + I_{rec} + I_{tun} + I_s} \qquad (1.23)$$

where I_p is the hole component of current due to injection of holes into the *n* emitter (the higher the doping level of the *n* emitter as compared to that of the *p* base, the smaller the share of I_p, see Fig. 1.18); I_{rec} is the current due to nonradiative recombination in the *pn* junction region; I_{tun} is the tunneling current formed by the carriers piercing across the potential barrier (this current grows with a decrease in the *pn* junction width and an increase in both the doping level of the base and forward voltage); and I_s is the leakage current over the *pn* junction surface.

When the injection ratio γ is equal to unity, the electron current I_n is equal to the total current I, so all the injected electrons contribute to the store of 'light energy' in the crystal.

The electrons injected into the *p* base recombine there near the *pn* junction. Along with emission of optical radiation (recombination centers *1* in Fig. 1.18) the processes of nonradiative recombination take place, of which the most important are the recombination via deep luminescent centers and Auger (collision) recombination.

1. Recombination via deep luminescent centers. An electron can indirectly pass to the valence band via recombination centers forming energy levels in the forbidden band (see the trapping level E_t in Fig. 1.12). In this case the energy of recombination is

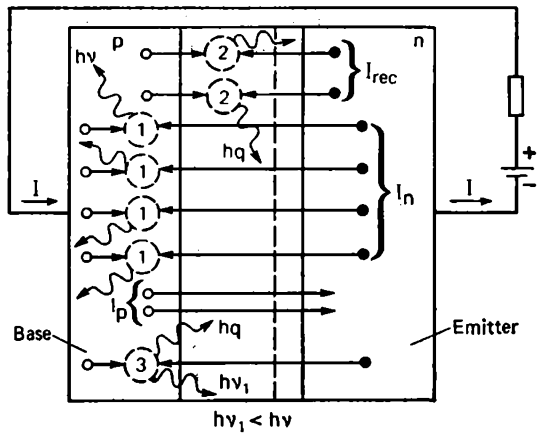


Fig. 1.18. Current components through a *pn* junction in injection electroluminescence

partially given out as long-wave photons and partially as phonons, i.e. thermal lattice vibrations (recombination centers *3* in Fig. 1.18). Both impurity levels and crystal defects can act as nonradiative recombination centers. The most harmful impurities in this respect are those which form energy levels near the middle of the bandgap (deep centers). To these impurities belong copper, nickel, cobalt, chromium, gold, and some other materials.

2. Auger recombination. When the concentration of free carriers in a semiconductor is very high, the probability of collision of three particles such as two electrons and a hole increases. The recombination energy of an electron-hole pair is then given up as a kinetic energy to a third free carrier. This kinetic energy is gradually lost during collision with the lattice.

From the practical viewpoint, the acts of recombination which cause the emission of photons whose energy is much lower than the gap energy E_g should also be regarded as nonradiative acts of recombination. The long wavelength radiation that results in this process extends beyond the operating spectral range of a radiator and is lost during transfer of the optical signal.

The effectiveness of radiative recombination is quantitatively defined by an internal quantum efficiency η_e, or yield, which is given as the ratio of the number of radiative recombination acts to the total number of acts of both radiative and nonradiative recombination. The internal quantum efficiency is sometimes expressed as the ratio of the generated photons to the number of minority carriers injected into the active region over the same period of time (in our case, these are electrons injected into the *p* base).

So, the product $\gamma\eta_e$ determines the effectiveness of injection elect-

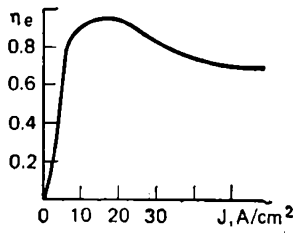


Fig. 1.19. Internal quantum efficiency plotted against forward current density

roluminescence. The upper limit of η_e depends on the conditions of manufacture of the *pn* junction and the electrical parameters specific to the operating conditions. The coefficient η_e is primarily a function of the forward current density J (Fig. 1.19). At low current densities, recombination in the space charge region exerts a strong effect on the value of η_e: it sharply rises with J until the diffusion component becomes the predominant component of the current through the diode. A further increase in J causes a gradual saturation of luminescent centers and a decrease in η_e.

The effect of temperature shows up as a change in both the injection ratio and internal quantum efficiency. The injection ratio somewhat grows at higher temperatures; the effectiveness of excitation of a radiator then commonly decreases, and so does too the internal quantum efficiency.

1.3. Light-Emitting Diodes

1.3.1. External quantum efficiency and radiation loss. Injection electroluminescence underlies the operation of light-emitting diodes (LEDs). The term LED also covers the diodes operating in the infrared (IR diodes).

The LED is the main and most universal incoherent radiator. Its advantages are the following: a high efficiency of conversion of the electric energy to optical energy; relatively narrow spectrum of radiation; ability to emit in almost any region of the optical range; comparatively narrow radiation pattern (included beam angle); low values of forward voltage drop, which enables electric compatibility with integrated circuits; high speed of response; small dimensions; processing compatibility with microelectronic devices; high reliability; and long life.

A figure of merit of a LED is the external quantum efficiency

$$\eta = \gamma \eta_e \eta_o \tag{1.24}$$

where γ is the injection ratio; η_e is the internal quantum efficiency; and η_o is the optical efficiency or light output efficiency which quantitatively defines the effectiveness of light extraction from the LED and is given as the ratio of the power of radiation emitted from the LED, P_{em}, to the power of radiation generated inside the crystal, P_{gen}:

$$\eta_o = P_{em}/P_{gen} \tag{1.25}$$

As noted in the preceding section, the product $\gamma\eta_e$ determines the efficiency of injection electroluminescence. But even if the value of $\gamma\eta_e$ is high, the external quantum efficiency of a LED may be low due to a decreased light extraction from the crystal bulk to the environment. The light extracted from the active region of a LED suffers losses due to the action of various mechanisms (Fig. 1.20).

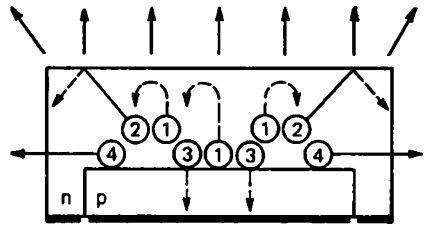


Fig. 1.20. Losses in light extraction from the active region

1. Self-absorption loss (beams *1*). The photons absorbed by a semiconductor can transfer their energy to valence electrons, thus exciting them to the conduction band. The photons can also give up their energy to free electrons in the conduction band or to holes in the valence band. This photon energy is also spent on the excitation of carriers to higher energy levels within a corresponding allowed band. Impurity absorption of photons is also possible, in which process the photons lose their energy by excitation of impurity levels. Besides, other types of absorption of photons in semiconductors can occur, for example, absorption by the crystal lattice, absorption attended by the transition of electrons from the acceptor to donor energy levels, etc.

2. Total internal reflection (TIR) loss. As the radiation impinges on the interface between two media differing in optical density, such as a semiconductor and air, a fraction of the radiation suffers total internal reflection (beams *2*). This radiation reflected back into the crystal is in essence lost through self-absorption.

TIR can heavily limit the external quantum efficiency. This effect is particularly strong in direct-bandgap semiconductors where almost the entire generated light subjected to TIR is lost by absorption. In indirect-bandgap semiconductors the internal absorption is much weaker, so the generated light has a greater probability of reaching some surface of the diode crystal. The loss in transmission of the light inside the diode structure is proportional to $V/S\chi_0$, where χ_0 is the absorption depth, V is the volume, and S is the total surface area of the LED crystal.

The light losses in both types of semiconductor are due to high indexes of refraction of the semiconductor materials used in LEDs ($n \approx 3.3$ to 3.8). These losses grow with a decrease in the bandgap energy. The radiation incident on the surface at an angle θ that is greater than the critical angle suffers total internal reflection. The radiation falling at an angle smaller than the critical angle undergoes partial reflection from the surface without an anti-reflection coating. The loss in both cases is known as the Fresnel loss. If

the surface of a semiconductor is given a dielectric coat of a certain thickness and specified index of refraction, the coat will produce an anti-reflection effect and thus raise the transmittance. But the critical angle practically remains unchanged.

3. Backscatter and edge radiation losses (rays *3* and *4*). Light generation in the active region of a semiconductor is spontaneous, so that the rays equiprobably propagate in all directions. The rays *3* advancing in the direction of the emitter are quickly absorbed. The active region often slightly differs in the index of refraction from the neighbor regions. For this reason the rays *4* merge together after multiple reflection and travel along the active region; hence, the edge radiation is stronger in intensity than the radiation escaping the crystal in other directions.

The external quantum efficiency η is thus the total index of the emissive power of a LED, which takes into account the injection ratio γ, internal quantum efficiency η_e, and optical efficiency η_o. To sum up, the external quantum efficiency is the ratio of emitted photons in a certain time to the number of electrons passing through the diode in the same time:

$$\eta = N_p/N_e \tag{1.26}$$

1.3.2. Heterojunction LEDs. The LEDs which use heterojunctions (heterostructures) feature the best emission parameters. Consider the main features of a heterojunction in more detail since it is the building block of many types of optoelectronic devices.

The heterojunction is a transition layer that appears at the boundary of two semiconductors differing in the energy gap. The semiconductors forming a heterojunction can have the same or different types of conductivity and so each pair of semiconductors can form four heterojunctions: p_1n_2; n_1n_2, n_1p_2, and p_1p_2. Here the index 1 relates to a wide-gap semiconductor, and the index 2 to a narrow-gap semiconductor.

In the heterojunction formed by semiconductors the charge carriers redistribute themselves and thus cause the contact potential difference to appear across the junction. The Fermi level for a heterostructure in the equilibrium state is common to both semiconductors. Fig. 1.21 displays the energy diagram of an emitting GaAlAs-GaAs heterostructure in the equilibrium state. At the metallurgical boundary of the junction, an energy gap appears, $\Delta E = E_{g1} - E_{g2}$. The heterostructure thus has different potential barriers for injected holes and electrons.

In the equilibrium state of the heterostructure only carriers of one type execute motion (electrons in the structure of Fig. 1.21). So, the forward bias applied to the structure causes a unidirectional injection of carriers—only electrons from the wide-gap layer (emitter)

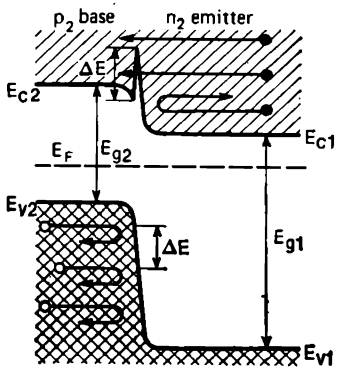


Fig. 1.21. Band diagram for a single heterostructure in equilibrium state

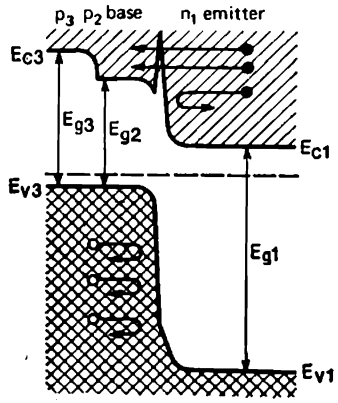


Fig. 1.22. Band diagram of a double heterostructure

to the narrow-gap layer (base). The structure that consists of a wide-gap emitter and a narrow-gap base is called a single heterostructure.

In distinction from a single heterostructure, a double heterostructure includes an additional wide-gap p_3 layer. This is the blocking layer whose conductivity type is the same as that of the base (Fig. 1.22). In the double heterostructure the second potential barrier inhibits the escape of electrons from the base region which forms a potential well for storing the injected electrons. The excess concentration of carriers in the active (emitting) region and the unidirectional injection sharply increase the internal quantum efficiency and response speed of the heterostructure.

Indeed, the double heterostructure ensures localization of injected carriers in the base of a width as small as a few micrometers. It is this feature that appreciably adds to the speed of response without reducing the internal quantum efficiency. In single heterostructures the emissive power sharply drops off as the base width diminishes, but the speed grows insignificantly. In the best single-heterojunction LEDs the external quantum efficiency is equal to 3 or 4% and the switching time t_{sw} ranges from 40 to 80 ns; in double-heterojunction LEDs the value of η is approximately the same, but the switching time drops to 20-30 ns.

It is of importance to stress the fact that, in contrast to conventional homogeneous junctions, the unidirectional injection in heterostructures is independent of the level of doping of the emitter and base regions and is kept stable up to high current densities. It is thus possible to vary the level of doping of heterostructure regions without impairing the injection capability of *pn* junctions.

One more feature of heterostructures is that the base and emitter differ in spectral properties. The spectral emission curve of the

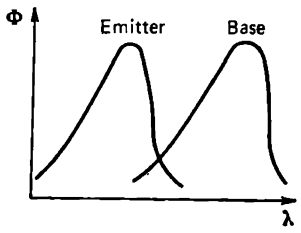


Fig. 1.23. Spectral characteristics of the base and emitter of the heterostructure

narrow-gap base proves shifted into the long wavelength region with respect to the spectral absorption curve of the wide-gap emitter (Fig. 1.23). For this reason the generated light escapes from the LED via the emitter practically without absorption.

Radiators using a double heterostructure whose active region lies away from the substrate are subject to the effect of multiple reflection (multipath effect). Being multiply reflected from various crystal faces as they experience TIR at the external boundary of the crystal, the beams at last fall on the external boundary at an angle that allows them to emerge outside. The multipath effect is obviously useful only if the amount of light absorbed in the process is small. Absorption in the narrow-gap base is somewhat offset by photoluminescence: absorption of a quantum leads to a new act of radiation.

All advantages of heterostructures are realizable only if the heterojunction produced fills the quality requirements. A quality heterojunction necessitates a good match between the parameters of the structure on both sides of the metallurgical boundary: the difference in lattice constants must not exceed 0.01% and the coefficients of thermal expansion must also be close in value. Where these requirements are not met, a high concentration of defects in the heterojunction region practically reduces to zero the advantages described above.

1.3.3. The LED as an element of the optron. Light emitting diodes find wide use as radiators in various data display systems, fiber-optic communication lines, etc., where they act as independent discrete optoelectronic devices. A LED can also serve as a constituent element of another optoelectronic device called an optron (optoisolator). In the latter case, the light emitting structure must ensure a high radiation power, the narrowest possible radiation pattern, and high speed of response. Only with this set of parameters can the radiator be a good match for the photodetector and ensure the optimal parameters of the optron. For ordinary LEDs used in other devices the requirements on the radiation pattern are commonly far less stringent. Thus LEDs used for visual applications may have a low response, i.e. a low speed of conversion of the electrical to optical energy.

Consider the typical structures and parameters of radiators (photoemitters) employed in modern optoisolators.

1. The structure with a *pn* junction formed in silicon-doped gallium arsenide (the so called mesa, see Ch. 5). The active region of this structure offers the following advantages: (a) almost complete ab-

sence of nonradiative recombination, indicative of a high structure perfection; (b) high injection efficiency; (c) difference in the spectral characteristics of radiation (peak wavelength $\lambda_{max} \approx 940$ nm) and absorption (cutoff wavelength $\lambda_c \approx 900$ nm), which sharply decreases self-absorption and provides a high optical efficiency.

In the best types of radiator based on this structure, the external quantum efficiency η is equal to 7-9% and the power conversion efficiency reaches 6 or 7%. Similar devices with a hemispherical active region have η ranging from 20 to 30% with the power conversion efficiency as high as 10-15%. However, a long lifetime of carriers and extended active region in these structures lead to the storage of a large excess charge in the forward bias diode, which takes 10^{-7}-10^{-6} s to be withdrawn by recombination and other mechanisms.

2. The single heterostructure based on a ternary compound GaAlAs. It ensures a unidirectional injection of holes into the base and efficient luminescence because the region of nonradiative recombination (defects at the substrate-base boundary) is slightly distant from the active region. Light extraction involves a minimum of absorption. In the best specimens of this structure, η is equal to 3 or 4% with the switching time t_{sw} ranging from 40 to 80 ns. The range of operating currents extends from 1 to 20 mA. The forward voltage drop is about 1.2 V at a forward current I_F of 10 mA. The breakdown voltage is equal to 6-8 V. The depletion (junction) capacitance under zero bias comes to 100 pF.

3. The double heterostructure based on the same ternary compound, GaAlAs. It allows for an increase in the concentration of an acceptor impurity, N_a, with the width of the base region kept invariable. The cutoff frequency f_c then grows, and so does too the tunnel component of forward current, while the external quantum efficiency declines. There is a certain optimum value of N_a at which the product ηf_c reaches a maximum. This product is the generalized index of performance of the radiator in the optron. As indicated earlier, the photoemitters of this type have η ranging from 2 to 2.5% and t_{sw} from 20 to 30 ns.

1.4. Parameters and Characteristics of LEDs

1.4.1. Emission characteristic. Depending on whether LEDs emit visible or invisible radiation, respectively described by the photometric or radiometric quantities, the receptors of this radiation can be either the human eye or any suitable photosensor. The largest usage of LEDs is in visual applications, which include digital indicators, sources of backlighting of legends and illumination of push

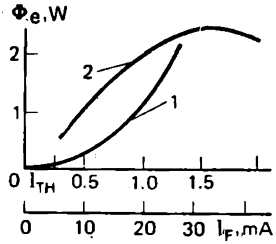


Fig. 1.24. LED emission characteristics (*1*) at small currents; (*2*) at high currents

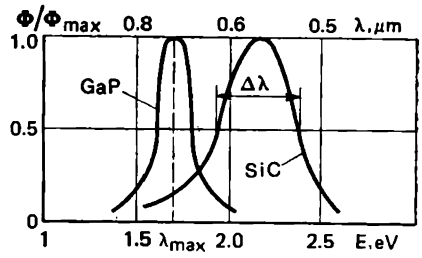


Fig. 1.25. Spectral characteristics of LEDs

buttons, electronic device status indicators, etc. The LEDs which provide the flux for detection by artificial means, namely the LEDs emitting infrared radiation (IR diodes) unperceived by the human eye, find use in tape/card readers, various communication and signaling systems, and digital devices.

The relative efficiency of LEDs is traditionally defined in terms of the light output versus the forward current through the LED and the emission wavelength. The emission characteristic of IR diodes relates the radiant flux Φ_e to the forward current (Fig. 1.24). For LEDs used for visual applications, the emission characteristic is usually given as the luminous intensity I_v plotted versus forward current I_F. The electric parameter used to define the LED performance is thus the forward current rather than the forward voltage. This is because the LED *pn* junction is normally driven on in the forward direction and the LED resistance is small, as is clear from the values of threshold current I_{TH} given in Fig. 1.24. The forward current through a LED can thus be regarded to be preset by the external circuit, adjusted in a wide range, and readily measured.

At low values of I_F, the recombination component of current is high and the injection ratio specified by expression (1.23) is low. As the forward current rises, the light output grows fast until the diffusion component becomes predominant.

A further increase in I_F leads to a gradual saturation of luminescent centers and a decrease in the emissive power of the LED. Besides, the probabilty of collision recombination increases too, which also diminishes the emissive power. The joint action of the above mechanisms that affect the radiant intensity results in a maximum appearing on the emission curve at a certain current. The maximum radiant (luminous) intensity depends on the area and geometry of the light emitting *pn* junction and on the size of electrical contacts.

1.4.2. Spectral characteristic. This characteristic is a plot of the parameters of radiation versus emission wavelength or the energy

of emitted photons. The wavelength of emission is a function of the difference in two energy levels between which electrons make radiative transitions. Since materials differ in band-gap energy, various types of LED emit at different wavelengths. The examples of the spectral characteristics of LEDs based on GaP and SiC with different impurities are given in Fig 1.25. Because during recombination of charge carriers the electrons execute transitions not between two energy levels, but between two groups of energy levels, the spectrum of radiation is found to be smeared. The spectral region of a LED is its radiation spectrum, $\Delta\lambda_{0.5}$, measured at half the height of the maximum of the curve. Most LEDs emit radiation that is close to quasimonochromatic radiation ($\Delta\lambda/\lambda_{max} \ll 1$) and exhibit a relatively narrow pattern of flux distribution in space.

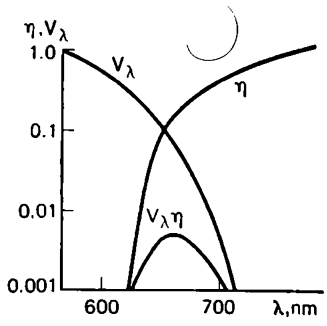


Fig. 1.26. Matching the spectral characteristic of LEDs to the luminous efficiency of the eye

No matter how efficient the LED is, the photodetector will fail to pick up the output light signal of even a high power if the emission wavelength does not correspond to the radiation spectrum to which the photodetector is responsive. In a majority of cases a LED must be spectrally consistent with the response of either the human eye or a silicon photodetector. The spectral response range of a photodetector extends from about 300 nm to 1 100 nm. The human eye responds to a substantially narrow range of wavelengths, which practically lies between 400 and 700 nm. For the operation of the radiator-detector pair to be effective, a thorough match must be secured between the spectral characteristics of both devices.

For example, to match a GaAsP LED to the human eye, we should select a wavelength at which the product of the spectral luminous efficiency V_λ for the eye and the LED quantum efficiency η is a maximum:

$$V_\lambda \eta_\lambda = \max \qquad (1.27)$$

As seen from Fig. 1.26, this maximum is reached at $\lambda = 655$ nm (the red color of luminescence).

In LEDs emitting at shorter wavelengths, for example, at $\lambda_{max} = 565$ nm (green color) and at $\lambda_{max} = 585$ nm (yellow color), the value of η is commonly much lower than that for red LEDs. But the relative response of the eye at the wavelength of red light is much higher.

As a result, it becomes possible to select a set of radiators, from the red to the green, which have the same value of the product $V_\lambda \eta_\lambda$ (with an accuracy of one order of magnitude).

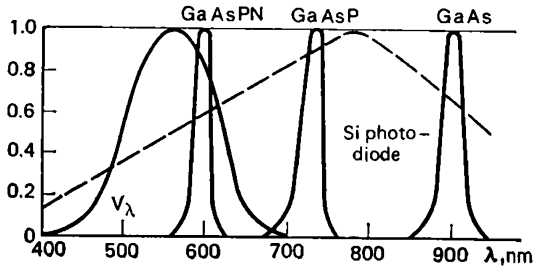


Fig. 1.27. Normalized spectral characteristics for the eye, V_λ, LEDs, and a silicon photodiode (dashed line)

Figure 1.27 displays for comparison the spectral characteristics of various LEDs and also the spectral response characteristics of the eye and a photodiode, the plots being made in normalized form. We should point to some features of spectral matching between a LED and a photodiode. On the one hand, this match is easier to secure as compared to that between the LED and the eye because the spectral range of the photodiode is wider. On the other, the spectral match is not always a decisive factor for the efficient operation of the LED-photodiode pair.

1.4.3. Optical parameters. These parameters describe the optical properties of LEDs. They enable the designer to select a LED of the desired luminescence color, which would give a maximum light output at the specified current. The optical parameters of a LED include an emission wavelength λ_{max}, dominant wavelength λ_d, luminous intensity I_v, and angle of radiation, θ (beam angle, or included off-axis angle).

The emission wavelength λ_{max} is the wavelength at which the spectral concentration (density) of the flux from a LED is a maximum. Modern structures have a narrow spread in the values of the emission wavelength for LEDs of the same type.

A LED intended for visual application is commonly fitted to the front panel of the device. It must be well seen in the ambient light. The level of general lighting is usually high. Therefore, for the visual perception to be more effective, it is necessary to increase the brightness contrast between the indicator and background. The contrast becomes more intensive with increased difference in luminescence between the on and off LEDs. A maximum difference can be achieved in two ways: by decreasing the reflection of ambient light from the LED surface and providing means for ensuring a maximum visual perception of the light emitted from the LED.

In the conditions of artificial lighting, contrast is increased with the aid of selective optical filters. The effect of this type of filter on the contrast range is evident from Fig. 1.28 which illustrates

a digital indicator as an example. It is readily seen that the filter appreciably improves the visual control of characters.

Fig. 1.28. Effect of a selective filter on display quality (without filter on the left and with filter on the right)

The main requirement placed on a contrast filter is that it must ensure good selectivity, i.e. transmit light only at the emission wavelength of the LED. The spectral characteristic of a filter is its relative transmission (transmittance) as a function of the wavelength of light The transmittance T_f of a filter is given by the relation

$$T_{f\lambda} = L_{v\lambda}/L_{v0} \tag{1.28}$$

where $L_{v\lambda}$ is the luminance at a wavelength λ with the filter; and L_{v0} is the luminance at λ_{max} without the filter.

Figure 1.29 illustrates how the transmittance T_f of a plastic filter varies with the emission wavelength at a filter material thickness of 1 mm (curve *1*) and 2.5 mm (curve *2*), with the index of refraction taken equal to 1.5. The value of T_f obviously depends on the refractive index n of the filter material. This dependence is well described by the expression of the form

$$T_{fn} = [2n/(n^2+1)]\,T_f \tag{1.29}$$

where T_f is the transmittance at $n = 1$ and wavelength λ; and T_{fn} is the transmittance at the same wavelength when $n \neq 1$.

In LED indicators used in combination with filters, the share of the luminous flux reaching an observer is proportional to the filter transmittance at the wavelength λ_{max} of the LED. Thus, let λ_{max} of the LED of interest be equal to 655 nm. Then T_f of the plastic filter 1 mm thick is equal to 0.8 when $n = 1.5$. This means that about 80% of the luminous flux emitted from the LED passes through the filter to the viewer, the remaining 20% of the flux being absorbed by the filter.

For LEDs used in nonvisual applications, the parameter λ_{max} determines the effectiveness of transmission of optical radiation from an IR diode to a photodetector, the share of light lost in the fiber-optic conductor, the choice of film sensitivity, etc.

The eye responds not only to the luminous intensity but also to the color of emitted light. Therefore, along with the peak wavelength λ_{max}, the dominant wavelength λ_d is used to define the color of a LED display. The parameter λ_d is a quantitative measure of color sensation induced in the eye by the display light. Two LEDs displaying different spectral characteristics will emit light of the same color if they have the same wavelength λ_d. The dominant wavelength is

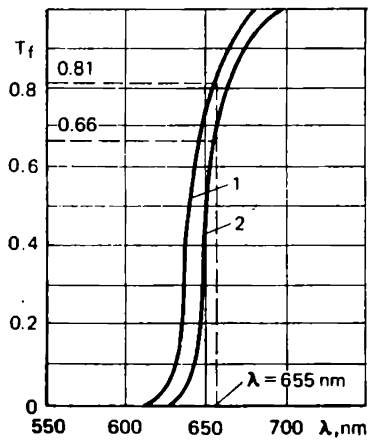


Fig. 1.29. Spectral characteristics of filters

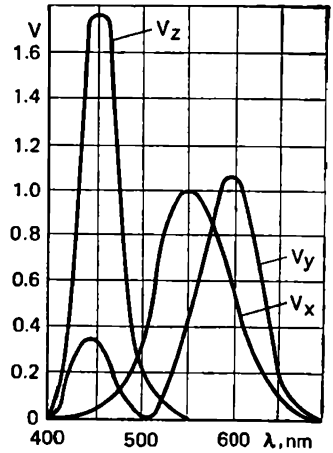


Fig. 1.30. CIE standard observer eye response curves

that wavelength of the light which, when mixed with the light of the CIE standard source C, evokes the same color sensation in the eye as that produced by the color spectrum from a LED. The CIE standard illuminant C is a 6 500 K color temperature source which simulates the daylight produced by an overcast sky.

The color of any light source can be estimated from the CIE color mixture curves V_x, V_y, and V_z. These curves represent the response of the eye's three receptors to the three primaries (reference stimuli)—the red, green, and blue (Fig. 1.30). It should be noted that the eye response curve V_y for the green color totally coincides with the spectral luminous efficiency function V_λ depicted in Fig. 1.2, which characterizes the general luminous response of the eye.

The color sensation evoked by the light from a radiator can quantitatively be defined in terms of chromaticity (color) coordinates. For these coordinates to be estimated, it is first necessary to plot the spectral characteristic of the radiator on the color response curves presented in Fig. 1.30 and then to calculate the tristimulus values defined as the amount of the three reference stimuli required to give a match with the light considered:

$$X = \int \frac{dP}{d\lambda} V_x \, d\lambda$$

$$Y = \int \frac{dP}{d\lambda} V_y \, d\lambda \qquad (1.30)$$

$$Z = \int \frac{dP}{d\lambda} V_z \, d\lambda$$

The above integrals are geometrically equal to the areas of the radiator's spectral characteristic plotted on the corresponding color mixture curves V_x, V_y, and V_z. Given the tristimulus values, expressed in relative units, we can now estimate the chromaticity coordinates

$$x = \frac{X}{X+Y+Z}$$
$$y = \frac{Y}{X+Y+Z}$$
$$z = \frac{Z}{X+Y+Z} \qquad (1.30a)$$
$$x+y+z=1$$

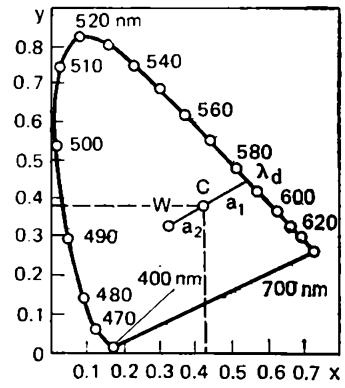


Fig. 1.31. Determining the dominant wavelength and color purity

These coordinates are used to plot a chromaticity diagram, commonly the plane diagram plotted in the rectangular x, y coordinates and employed to specify the color of a light source through the primary colors (Fig. 1.31). The solid horseshoe-shaped curve on this graph is the locus of points representing the chromaticities of spectrally pure stimuli and is known as the spectrum locus. The point of white color, W, that results from the blending of the three primaries of equal luminance lies in the center of the chromaticity diagram (the coordinates $x = 1/3$ and $y = 1/3$).

To determine the dominant wavelength of a light source, the x, y color coordinates are plotted on the CIE chromaticity diagram (the x, y point C in Fig. 1.31). A line is then drawn from the point W through the point C until it intersects the line of pure colors. The point where the line intersects the perimeter of the diagram is the dominant wavelength λ_d of the source. This point also defines the color of the source, which is yellowish-orange, as is clear from the chromaticity diagram of Fig. 1.32.

The color purity, or saturation m, is defined as the ratio of the segment a_1 to the sum of segments a_1 and a_2:

$$m = \frac{a_1}{a_1+a_2} \qquad (1.31)$$

The chromaticity coordinates for most LEDs plot very close to the perimeter of the chromaticity diagram, so the color purity approaches a value of 1.

Good visual perception of the colors of multicolor indicators presupposes adequate color discrimination. Multicolor indicators usually provide an optical indication of operating conditions, for example, the turn-on and turn-off states of devices. The spectral characteristics

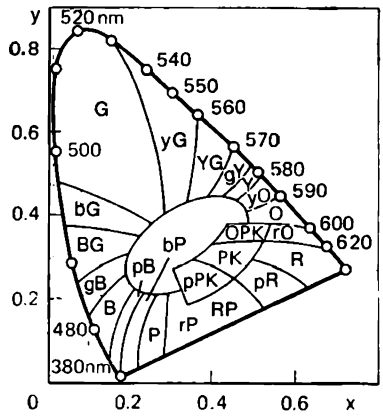


Fig. 1.32. Dominant wavelengths and corresponding colors shown on the CIE chromaticity diagram
R — red, *rO* — reddish orange, *O* — orange, *yO* — yellowish orange, *Y* — yellow, *gY* — greenish yellow, *YG*— yellow green, *yG* — yellowish green, *G* — green, *bG* — bluish green, *BG* — blue green, *gB* — greenish blue, *B* — blue, *pB* — purplish blue, *bP* — bluish purple, *P* — purple, *rP* — reddish purple, *RP* — red purple, *pR* — purplish red, *pPK* — purplish pink, *PK* — pink, *OPK* — orange pink

for six main colors discriminated under low and high illumination appear in Fig. 1.33. These characteristics were plotted in the following way: the dominant wavelength of the light incident on the surface was varied in discrete steps, each 4 to 7 nm long. A large number of observers then identified the surface color after each successive step. Figure 1.33 also displays the dominant wavelengths for typical LEDs. The resultant curves allow us to conclude that the blue, green, and red are well discriminated colors which encompass wide bands of the spectrum. Along with these three primaries, the orange or yellowish-orange can be used as a complementary (fourth) color.

Reference books and manuals usually cite the values of both λ_{max} and λ_d. Knowing λ_d, we can determine the actual color produced by a LED from the CIE chromaticity diagram of Fig. 1.32. Thus the yellow LED lamp proves yellowish-orange (λ_d = 585 nm) and the green LED is actually greenish-yellow. A contrast filter transmits at a minimum of loss if its color is the same as that of the LED.

The luminous intensity I_v is the common parameter that quantitatively describes the radiation from a LED. But this parameter varies with the direction of radiation. The off-axis intensity of

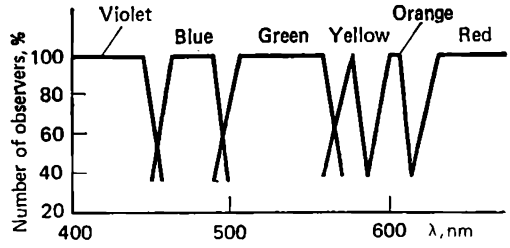


Fig. 1.33. Spectral characteristics illustrative of color discrimination

a device can be estimated from the radiation pattern which specifies the included off-axis beam angle θ where the luminous intensity is equal to one half of the axial intensity (Fig. 1.34).

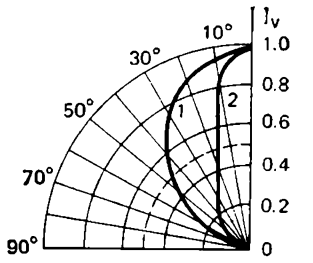


Fig. 1.34. LED radiation patterns

Using the radiation pattern, we can estimate the luminous flux perceived by the eye, when viewed at a certain angle to the geometric axis of the radiator. The rated radiation pattern enables us to evaluate the luminous intensity at any viewing angle. As an example, Fig. 1.34 illustrates the radiation patterns (curves *1* and *2*) for two different LEDs. The beam angles for the first and the second pattern are equal to about 45° and 15° respectively.

1.4.4. Electric parameters of LEDs. The parameters of a LED as an element of a dc circuit follow from its voltage-current (*VI*) characteristic (Fig. 1.35). Variations in the forward *VI* characteristics of LEDs are due to the difference in energy gaps of the materials used. The shorter the wavelength of emission, the higher the forward voltage drop across the LED and the greater the power dissipation in the diode. The reverse *VI* characteristics show small permissible values of the reverse voltage because the *pn* junction width in a LED is insignificant. When using a LED for work in the circuit exhibiting large reverse voltages, a normally reverse biased diode must be placed in series with the LED to guard it against breakdown.

The switching speed of a LED is a function of the time lag of the emission process on applying a dc step input of the rectangular form (Fig. 1.36). The switching time is the sum of the turn-on time t_{on} and turn-off time t_{off} of emission. The time lag of a LED depends on the process of recharge of the depletion capacitance (*pn* junction capacitance) and the processes of storage and withdrawal of minority carriers in the active region of the LED. For LEDs used in visual applications, the speed of response is of secondary importance since the persistence of the human eye reaches about 50 ms, which is much higher, than the switching time t_{sw} for a LED. For the systems of data recording and retrieval without visualization, such as optoisolators, the switching time of an IR diode is part of the total switching time of the optoelectronic device. It is certainly desirable that t_{sw} of the IR diode should be as small as possible.

LEDs are prone to degradation, or aging, which appears as a decrease in the light output with time. Degradation is due to an increase in the number of nonradiative recombination centers brought about by uncontrollable impurity atoms traveling in the electric field.

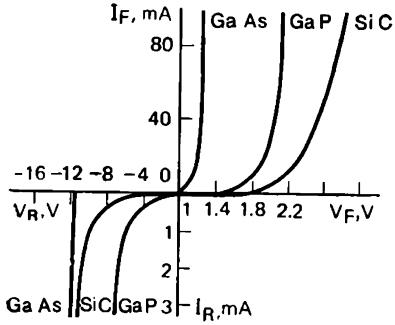


Fig. 1.35. LED *VI* characteristics

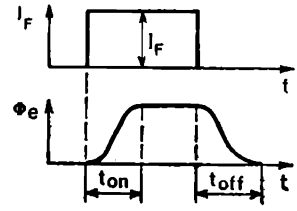


Fig. 1.36. LED response to unit step input

One more cause is the deactivation of some radiative centers as they move from lattice sites into interstices.

A decrease in the luminous flux due to degradation obeys the exponential law:

$$\Phi_v(t) = \Phi_v(0) \exp(-t/t_{dgr}) \quad (1.32)$$

where $\Phi_v(0)$ is the initial flux generated by a LED; and t_{dgr} is the time constant accounting for the rate of degradation.

The service life of most LEDs on the average reaches 10^4 hours and can even be as long as 10^5-10^6 hours, though for some types it does not exceed 10^3 hours.

Consider now in brief how temperature affects the performance of LEDs.

The peak wavelength λ_{max} generally rises somewhat with temperature because of a decrease in the energy gap of semiconductors under heat. The radiant (luminous) flux decreases with growing temperature. In a LED used for visual indication, the luminous intensity diminishes with increasing temperature because of a decrease in both the emissive power and the eye sensitivity with a change of the wavelength. For example, the eye sensitivity to the light of a red LED emitting at 650 nm varies by about 4.3% with a change of the wavelength by 1 mm. The wavelength shift reaches 0.2 nm°C^{-1}. The luminous intensity drops by about 1% with a temperature increase by 1°C. The total relative variation of luminous intensity with temperature growth then amounts to

$$\Delta I_v/\Delta T = -4.3 \times 0.2 - 1 = -1.86\%\ °C^{-1} \quad (1.33)$$

The luminous intensity of a LED varies with temperature in an exponential manner

$$I_v(T) = I_{v0} \exp[-k(T - T_0)] \quad (1.34)$$

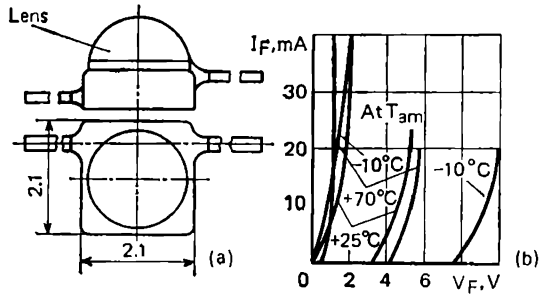


Fig. 1.37. Silicon carbide LED
(*a*) external view; (*b*) boundary positions of forward *VI* curves

where I_v (T) is the luminous intensity at temperature T; I_{v0} is the luminous intensity at room temperature; and $k = \ln \mathrm{TC} I_v$ (TC is the temperature coefficient of I_v).

1.4.5. Parameters and characteristics of industrial LEDs. In conclusion of this section let us look into the typical parameters and characteristics of commercially available LEDs.

Silicon-carbide diffused LEDs (Fig. 1.37) are yellow lamps intended for work in various devices. A typical encapsulation for a LED is a plastic package with flat, flexible, tinplated leads and a glass lens. The mass of one LED is not over 0.5 g.

The values of the main parameters at an ambient temperature T_{am} of 25 $\pm$ 10°C are the following: the luminance ranges from 10 to 20 cd m^{-2} at maximum permissible direct forward currents I_F of up to 40 mA; the forward voltage V_F at the above values of current is not over 5.5 V.

These devices show an appreciable spread in characteristics from type to type and a strong temperature dependence of the forward *VI* curves. In Fig. 1.37*b* are shown the boundary positions of the forward branches of the curves at different ambient temperatures.

LEDs normally operate at forward bias. The reverse voltage surges must not exceed 3 V. The luminance from these diodes is also heavily dependent on temperature.

Epitaxial gallium phosphide LEDs are red and green lamps designed for operation in various electronic devices. The external view of a GaP diode is shown in Fig. 1.38*a*; Fig. 1.38*b* illustrates how the luminance of these diodes varies with I_F. Each diode weighing about 0.25 g is fitted into a metal envelope furnished with tinplated wire leads and a glass lens.

In operation at $T_{am} = 25 \pm 10$°C the forward voltage V_F does not exceed 5 V at forward currents of 5 to 30 mA. The luminance at these values of I_F lies between 5 and 50 cd m^{-2}. The range of $I_{F\,max}$ extends from 10 to 30 mA for different types operating at

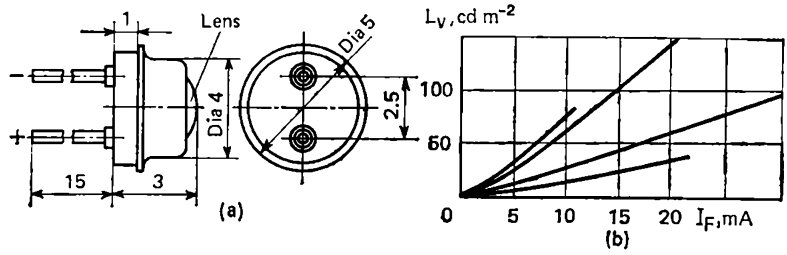


Fig. 1.38. Gallium arsenide LED
(*a*) external view; (*b*) emission characteristics

ambient temperatures from plus 70°C to minus 60°C. The reverse voltage surges are set at 6 V.

Gallium arsenide (GaAs) diffused mesa IR diodes find use in electronic equipment designed for a variety of purposes. Each diode weighs 0.5 g. The encapsulation is a metal envelope fitted with tin-plated wire leads and a glass lens. Various types of IR diode can produce a radiant flux of up to 0.6 mW at a direct forward current of 100 mA.

The values of electric parameters at an ambient temperature T_{am} of 25 $\pm$ 10°C are the following: the turn-on time 30 ns; forward voltage not above 1.7 V at I_F = 100 mA; maximum reverse voltage 4 V (diodes normally operate at forward bias); maximum direct forward current 120 mA. The range of limiting ambient temperatures extends from plus 85°C to minus 60°C.

IR diodes exhibit a spread in the values of their characteristics and a strong temperature dependence. The dash lines in Fig. 1.39*a* illustrate the extreme possible positions of forward *VI* characteristics at 25°C.

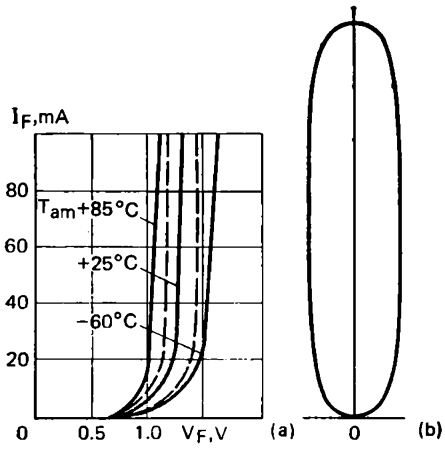


Fig. 1.39. IR diode
(*a*) forward *VI* curves; (*b*) radiation pattern

A typical radiation pattern for IR diodes appears in Fig. 1.39*b*. The included angle reaches 20°; the possible positions of the optical axis for 95% diodes lie within an angle of 10°.

At present industry produces in quantities semiconductor indicators which present information in the form of characters such as digits and letters. These indicators are known as *character displays*, each consisting of a certain number of electroluminescent elements. The display containing eight light-emitting ele-

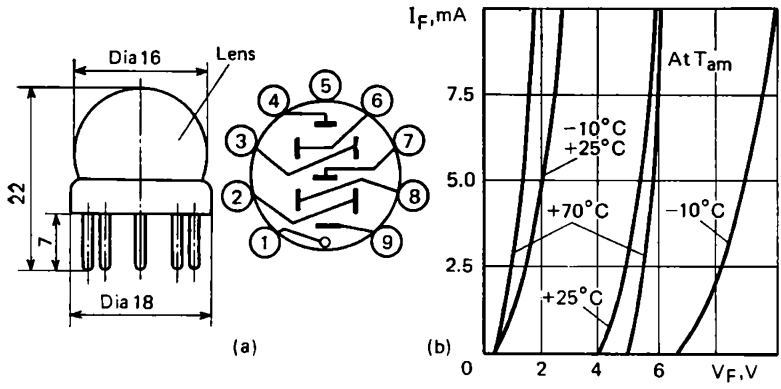


Fig. 1.40. Character display
(*a*) external view; (*b*) boundary positions of forward *VI* curves

ments is able to reproduce all digits from 0 to 9 and some letters (altogether 27 characters) and also the sign of the character position. General data on character displays are given in Table 1.3.

Table 1.3. Some Parameters of Character Displays

Parameter	Value
Operating current, mA	40
Included angle, deg	150
Character height, mm	9
Switching time, μs	1
Service life, h	10 000

Character displays are most advantageous for use in small-size equipment, readers of digital information from computers, control systems, and digital measuring devices.

One such device is a silicon carbide (SiC) character indicator illustrated in Fig. 1.40*a*. It is made in IC form and intended for work in equipment performing a variety of functions. The indicator produces the yellow color and can display all the numerals from 0 to 9. The metal envelope of the indicator is fitted with a glass lens and a 9-pin frame for connection to the voltage source. The device weighs no more than 7 g.

The included off-axis angle is equal to 60°. The indicators of this type operate in the temperature range from minus 10°C to plus 70°C. The devices have an appreciable spread in characteristics and strongly depend on temperature. Fig. 1.40*b* displays the extreme positions of forward *VI* curves for 95% devices of this class at different ambient temperatures.

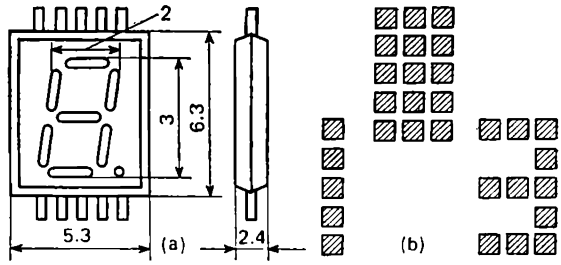


Fig. 1.41. Typical dimensions of a semiconductor seven-segment display (*a*) and display screens (*b*)

Typical dimensions of character displays composed of LEDs are shown in Fig. 1.41. Seven-segment (eight-element) displays (Fig. 1.41*a*) and 3×5-element matrix displays (Fig. 1.41*b*) commonly find use in numeric display applications. For indicators intended to display both numerals and letters, the arrays of 5 by 7 LEDs are more convenient.

The size of a display varies with distance from a viewer. The height of characters is so chosen that the viewing angle should be equal to 10-24′. The viewing angle β is given by the expression

$$\beta = 120 \arctan (h/2d) \qquad (1.35)$$

where h is the character height; and d is the distance from the indicator to the eye.

Small-size numerical displays enjoy wide uses in wrist devices such as watches and in portable calculators. The typical height of characters reaches 2.5 to 3.5 mm, which corresponds to a viewing angle of 9.5-13.4′ for a distance equal to the length of an extended arm.

Other indicators with typical character heights of 8 to 12 mm find use in device panels or control desks. At a distance of 3 m, the viewing angle is equal to 9.2-13.8′. The character width-to-height ratio for the two types of display commonly lies between 0.6 and 0.8.

Chapter 2

Photodetectors

Photodetectors are photo-electronic devices which convert the energy of optical radiation to electric energy. These devices rely on the internal photoelectric effect observed in semiconductor materials. In certain materials, both solids and liquids, the energy absorbed

from the radiation can be sufficient to emit the freed electrons from the material. This is the external photoelectric effect, called the photoemissive effect and used in such devices as photoemissive tubes, photomultipliers, and image intensifiers.

The features of photo-electronic devices that determine their successful use in many applications are a high efficiency of converting the energy of photons to the energy of mobile carriers of charge, internal amplification of the converted energy, high speed of response, and wide functional capabilities. Of all the variety of photoelectronic devices such as photodiodes, phototransistors, photothyristors, and photoresistors, it is photodiodes that most fully comply with the requirements for the characteristics mentioned above. Although photodiodes are not capable of amplification, this drawback is rather conditional. The thing is that at the modern level of integrated technology, it is much easier to introduce amplification means in the channel of electric signal transmission than to ensure internal amplification in the photodetector proper. The modern status of optoelectronics appreciably depends on the progress in the area of photodiode structures. Therefore, in this chapter the focus of attention will be on photodiodes.

2.1. Internal Photoelectric Effect

2.1.1. Photoconduction and photo-emf. Photodetectors are light-activated semiconductor devices. The radiation falling on a semiconductor material changes its electrophysical parameters as it produces additional free carriers and thus increases the conductivity (photoconductivity) of the material. This is the internal *photoeffect*. Photodetectors use two forms of this phenomenon: the photovoltaic effect and photoconductive effect.

The photovoltaic effect appears in semiconductors with an internal potential barrier (across the *pn* junction, metal-semiconductor junction, or heterojunction): the internal electric field of the depletion layer at the junction separates the photocarriers (electrons and holes) generated by the incident radiation. The spatially separated electrons and holes produce a photoelectromotive force (photo-emf).

The photoconductive effect (photoconduction) shows itself as the generation of photocarriers and thus an increase in the conductivity of the semiconductor when exposed to radiation.

Both effects are taken advantage of in practical applications: the photovoltaic effect is used in photodiodes, phototransistors, photothyristors, and other *pn* junction devices; the photoconductive effect is used in photoresistors.

Consider the process of generation of additional holes and electrons in a semiconductor as it absorbs the energy of photons (Fig. 2.1).

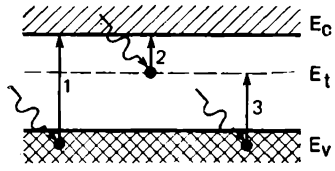


Fig. 2.1. Intrinsic (transition *1*) and extrinsic (transitions *2*, *3*) photoconductive effects in semiconductors

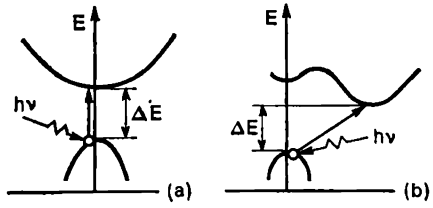


Fig. 2.2. Electron transitions from the valence to the conduction band due to the intrinsic photoconductive effect (*a*) direct transition; (*b*) indirect transition

The photons can give up energy to electrons in the valence band to enable them to jump to the conduction band; in other words, the photons transfer energy to electrons and thus ionize semiconductor atoms. The process of generation of intrinsic photocarriers which execute direct transitions from the valence to the conduction band (transition *1*) is known as the intrinsic photoconductive effect.

Impurity absorption is also possible, in which process the energy of photons is spent on the ionization and excitation of impurity atoms (transitions *2* and *3*). This is the extrinsic photoconductive effect. The concentration of impurity atoms is small, and these atoms become ionized even at relatively low temperatures. So, the intrinsic photoconductivity is substantially higher than the extrinsic (impurity) photoconductivity—the intrinsic photocarriers account for a major share of carriers.

In the process of intrinsic absorption of photons, valence electrons can make direct transitions to the conduction band, i.e. without a change in their momenta (Fig. 2.2*a*, see also Sec. 1.1). Valence electrons can also move to the conduction band by making indirect transitions which entail changes in their momenta (Fig. 2.2*b*). For indirect transitions to be made, a third quasiparticle must take part in the absorption process in addition to the photon and electron, according to the law of conservation of momentum. This quasiparticle is a quantum of thermal energy of the semiconductor crystal lattice.

For an incident photon to be able to produce a photocarrier, the following energy relations must obviously hold (see Fig. 2.1):

$$\begin{aligned} E_{P1} &= h\nu_1 \geqslant E_c - E_v \\ E_{P2} &= h\nu_2 \geqslant E_c - E_t \end{aligned} \tag{2.1}$$

where E_{P1} and E_{P2} are the energies of photons; E_c, E_v, and E_t are the energy levels at the bottom of the conduction band, the top of the valence band, and the impurity trap center respectively.

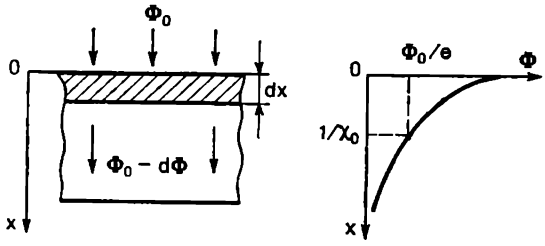


Fig. 2.3. Light absorption in a semiconductor

So, the intrinsic photoeffect can take place in a semiconductor only if the incident radiation has a wavelength that is equal to or smaller than a certain critical value:

$$\lambda_c = hc_0/(E_c - E_v) \approx 1.23/E_g \qquad (2.2)$$

where E_g is the bandgap energy, eV; and c_0 is the speed of light in a vacuum.

The critical wavelength λ_c, μm, is the longest wavelength that specifies the long-wavelength cutoff of the spectral response of a material. At this limiting wavelength photons still contain sufficient energy to produce free carriers within the material. At $\lambda > \lambda_c$, the intrinsic photoeffect is impossible; at $\lambda < \lambda_c$ this effect may take place. In photoconductive detectors based on silicon, germanium, gallium arsenide, cadmium sulfide, and cadmium selenide, the critical wavelength λ_c is equal to 1.1, 1.8, 0.9, 0.7, and 0.8 μm respectively.

2.1.2. Photoeffect parameters. A measure of the effectiveness of photoelectric processes is the quantum yield, or quantum efficiency η_P, which is equal to the ratio of generated electron-hole pairs to the number of absorbed photons. In the operating spectral range of photodetectors, the quantum efficiency is usually equal to unity, i.e. each absorbed photon produces an electron-hole pair.

The absorption of radiation in semiconductors is defined by a penetration (absorption) depth χ_0 or by the reciprocal quantity $1/\chi_0$, called the absorption coefficient, which specifies a relative change of the radiant flux in a semiconductor layer (Fig. 2.3):

$$\frac{1}{\chi_0} = -\frac{1}{\Phi(x)}\frac{d\Phi}{dx} \qquad (2.3)$$

where $\Phi(x)$ is the radiant flux penetrating a distance x into the semiconductor bulk.

Dividing the variables in expression (2.3), we can readily obtain Bouguer's law which describes the radiant power attenuation in a

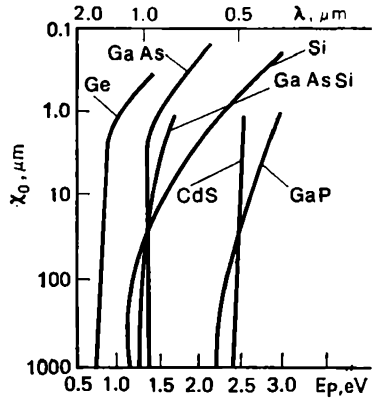


Fig. 2.4. Spectral characteristics of light absorption in photodetector materials

semiconductor due to absorption:

$$\int_{\Phi_0}^{\Phi} \frac{d\Phi}{\Phi} = -\frac{1}{\chi_0} \int_0^x dx;$$

$$\Phi(x) = \Phi_0 \exp(-x/\chi_0) \qquad (2.4)$$

where Φ_0 is the radiant flux at the surface of a semiconductor; and χ_0 is the penetration depth.

Thus the penetration depth χ_0 can be defined as the thickness of a semiconductor layer where the radiant flux decays exponentially as it passes through this layer and falls to $1/e$ of its value at the surface ($e = 2.718$). The relation between the penetration depth and the photon energy or emission wavelength specifies the absorption spectrum of a semiconductor, or the spectral characteristic of absorption (Fig. 2.4). It should be pointed out that the penetration depth χ_0 for most photoconductive materials, excepting silicon, varies quite abruptly near the critical wavelength λ_c. Silicon, however, passes from the fully transparent state ($\chi_0 \to \infty$) to the fully nontransparent state, in which it completely absorbs radiation at the surface ($\chi_0 \to 0$), when the wavelength λ changes by a factor of about 2.

2.2. Parameters and Characteristics of Photodetectors

A certain set of figures of merit are used to estimate and compare the performance of different types of radiation detector, whatever their physical nature and design.

2.2.1. Responsivity. This basic parameter can to a certain extent be considered analogous to the gain of *n*-type devices. In the general case, the responsivity of a photodetector reflects a variation in the electrical output with a unit input of incident radiation. In other words, the responsivity of a detector is the ratio of a change in the measured electrical quantity to the magnitude of radiant energy incident on the detector (response per unit radiation input).

Considering that optical radiation is definable through both radiometric and photometric quantities, as listed in Table 1.1, we can naturally identify the radiant flux responsivity $S_{\Phi e}$, luminous flux responsivity $S_{\Phi v}$, irradiance responsivity S_{Ee}, and illuminance (illumination) responsivity S_{Ev}.

Depending on the electrical parameter measured at the detector output, we can differentiate between the current responsivity S_I

and voltage responsivity S_V. In the first case, the measured quantity is photocurrent I_P and in the second it is photovoltage V_P of an output signal. The expressions for S_I and S_V are written as

$$\begin{aligned} S_{I\Phi v} &= I_P/\Phi_v; & S_{IEv} &= I_P/E_v \\ S_{V\Phi e} &= V_P/\Phi_e; & S_{VEe} &= V_P/E_e \end{aligned} \tag{2.5}$$

where $S_{I\Phi v}$ and S_{IEv} are the current responsivities to the luminous flux and illuminance respectively; $S_{V\Phi e}$ and S_{VEe} are the voltage responsivities to the radiant flux and irradiance respectively.

The responsivity of a photodetector generally varies with the radiation input. The notions of static and differential responsivities are used to account for this variation. The static responsivity is given as the ratio between the steady-state values of the output and input quantities. The differential responsivity is the ratio between the small increments of these quantities. For example, the differential current responsivity to differential illuminance is

$$S_{IEv} = \Delta I_P/\Delta E_v \tag{2.6}$$

A photodetector responds differently to different wavelengths of incident radiation. So, one has to discriminate between the total (integral) responsivity and monochromatic responsivity. The first quantity is the response to the polychromatic radiation of the given spectrum, and the second quantity is the response to the monochromatic radiation of wavelength λ.

2.2.2. Noise and threshold parameters. In addition to the desired useful signal, a spurious signal of random amplitude and frequency is always present at the detector output. This random signal is the detector noise. The sources of noise can be both internal and external to the photodetector. Noise does not allow the detector to pick up vanishingly small optical signals since they become indistinguishable against the noise background. In an adequately designed photodetector the response to small input signals only depends on the level of intrinsic noise in the device. Noise arises from random (fluctuation) processes, so the level of noise is defined by probabilistic parameters such as expectation (mean noise level) and variance (mean-square value, or dispersion). The distribution of noise power over the spectrum is often given as the spectral noise density—noise per unit frequency band.

Along with thermal and shot types of noise common to semiconductors, photodetectors also suffer from radiation (photon) noise which appears as fluctuations of an optical signal picked up by a photodetector, i.e. fluctuations of photons falling on the photosensitive layer.

The noise of a photodetector is commonly expressed in terms of noise current or noise voltage. The noise current I_N is a mean-square fluctuation of current passing through a detector in the specified

frequency band. The noise voltage is a mean-square fluctuation of voltage across the given load resistance in the detector circuit.

One more figure of merit, often used to relate the responsivity of a detector to its noise is known as the noise equivalent power (NEP), or the threshold flux that gives a signal equal to the noise in a bandwidth of 1 Hz. More precisely, this is the rms value of the radiant flux acting on the photodetector that produces the rms value of photocurrent equal to the rms values of noise. Thus if $\Phi = \Phi_{NEP}$, then the noise current is equal to the photocurrent, $I_N = I_P$.

Since noise depends on the frequency of the measurement, Φ_{NEP} is a function of the frequency. Therefore, it is customary to specify the NEP with respect to a unit frequency band. The unit bandwidth NEP, Φ_{NEP1}, is a minimum rms value of the sinusoidally modulated flux of the given spectrum normalized to a unit bandwidth and expressed in $WHz^{-1/2}$. The level of photodetector noise depends on the area of the sensitive element. To account for this dependence, the specific parameter is used to relate Φ_{NEP1} to the area:

$$\Phi^*_{NEP\,1} = \Phi_{NEP\,1}/A \tag{2.7}$$

where Φ_{NEP1} is the unit bandwidth NEP; and A is the area of the sensitive element of the photodetector. The specific NEP is thus the threshold flux in a unit bandwidth, related to a unit area of the sensitive element. The quantity that is the reciprocal of Φ^*_{NEP1} is called the area weighted detectivity D^*, expressed in cm $W^{-1}Hz^{1/2}$:

$$D^* = A/\Phi_{NEP\,1}{}^{1}$$

2.2.3. Basic characteristics. These include the voltage-current, spectral, and energy characteristics. The *VI* characteristic relates the output voltage of a detector to its output current (photocurrent) at the specified value of the radiant flux. The spectral characteristic is the spectral response of a photodetector to the wavelength of the incident monochromatic flux. The energy characteristic describes the photocurrent as a function of the radiant flux falling on the detector. The energy characteristic (Fig. 2.5) usually takes the form of a power function

$$I_P \sim \Phi^n \tag{2.8}$$

The exponent n defines the linearity of the energy curve. When $n \approx 1$, the photodetector responds linearly. The range of Φ from Φ_{min} to Φ_{max}, over which the characteristic is linear, specifies the dynamic range $\Delta\Phi$ of the linear response, commonly expressed in decibels:

$$\Delta\Phi = 10 \log \frac{\Phi_{max}}{\Phi_{min}} \tag{2.9}$$

A typical spectral response of a photodetector is shown in Fig. 2.6. The critical wavelength λ_c defined by expression (2.2) re-

[1] Translator's note.

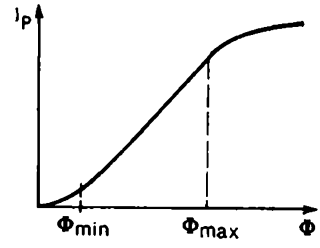


Fig. 2.5. Photocurrent plotted against the flux incident on a photodetector

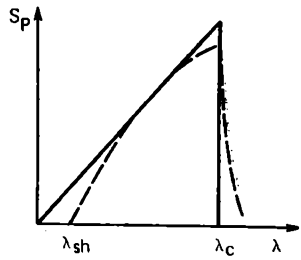


Fig. 2.6. Spectral response of a photodetector

presents the long-wavelength cutoff of the spectral response of a photodetector to the incident radiation. The other cutoff wavelength, the shortest wavelength λ_{sh}, specifies the short-wavelength cutoff, beyond which the material heavily absorbs radiation.

2.2.4. The photodetector as an element of the optron. The function of the photodetector here is to pick up optical signals from a radiator, most often an IR diode. Since radiators employed in optrons have to emit within a relatively narrow spectrum, the specific shape of the spectral characteristic for the photodetector is of no importance and its integral parameters over the spectrum have no meaning. What is important is a maximum response of the detector at the operating wavelength of the incoming radiation.

The speed of a photodetector depends on its switching time t_{sw}. In modern detectors of the main classes the switching time lies within 10^{-5}-10^{-7} s, and high-speed detectors have as small values of t_{sw} as 10^{-8}-10^{-10} s; the progress in optoelectronics calls for yet faster detectors with t_{sw} covering the range of 10^{-10} to 10^{-12} s. A high speed is not realizable where the load has a high resistance since the photodetector capacitance then takes much time to charge and the switching time becomes long. On the other hand, the detector of radiant fluxes close to Φ_{NEP} in values calls for a load resistance as high as 10^7 ohms; the time it takes the detector capacitance (C = 5 to 20 mF) to charge then becomes as long as 10^{-4}-10^{-5} s. It is thus practically impossible to achieve a high speed when the photodetector operates in the 'threshold mode', i.e. when $\Phi = \Phi_{NEP}$ and I_P is small. The threshold parameters of photodetectors used in optrons are of minor importance.

2.3. Photodiodes

2.3.1. The principle of action. The photodiode is a photosensitive semiconductor diode relying on the photovoltaic effect. Fig. 2.7 demonstrates a simplified structure of the photodiode.

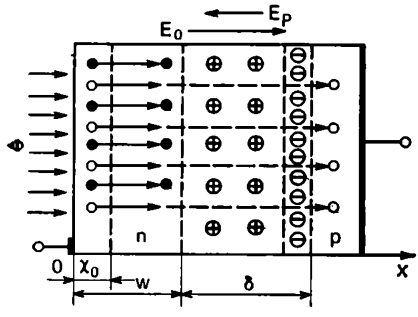


Fig. 2.7. Photodiode structure on exposure to radiation

Let the radiation act in the direction normal to the *pn* junction plane. As the material absorbs photons of energy higher than the bandgap energy, electron-hole pairs (photocarriers) appear in the *n*-type base at a depth of about χ_0 from the semiconductor surface. The photocarriers diffuse into the *n* region. The width w of the *n* region is such that a major share of photocarriers have no time to recombine and reach the *pn* junction boundary ($x = w$), where the electric field E_0 separates the electrons and holes. The holes move into the *p* region, while the electrons cannot negotiate the junction field E_0 and pile up at the *pn* junction boundary in the *n* region. The current of photocarriers through the *pn* junction thus results from the drift of minority carriers—holes.

The magnitude of photocurrent through a photodiode depends on the level of illumination. It is the optical control of the current through a photodiode that determines the specific operating conditions of this photodetector. To illustrate the point, compare the principle of a normal *pn* junction diode with that of a photodiode.

In the equilibrium state when there is no illumination of the diode element, the photodiode acts as a normal junction diode, and the former has a band diagram just like that of the latter (Fig. 2.8*a*). The Fermi level E_F is the same in both regions, and the diffusion component of current through the *pn* junction is equal to the drift component due to minority carriers. The forward voltage applied to the *pn* junction of the normal diode (with the *p* region on the positive side and *n* region on the negative side of the source) disturbs the equilibrium state as it lowers the internal potential barrier. Of

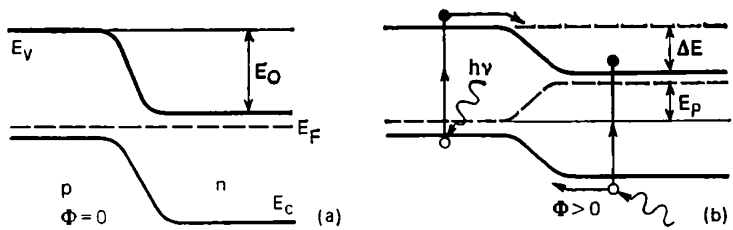


Fig. 2.8. Internal photoeffect in a *pn* junction
(*a*) *pn* junction in equilibrium state; (*b*) *pn* junction on exposure to radiation

the two flows of carriers traversing the junction, the diffusion flow contributes more to the current and becomes predominant at a sufficiently high forward bias. The drift component of current through a normal junction diode is in fact a parasitic component and must be as small as possible.

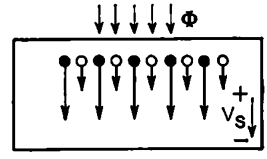


Fig. 2.9. Generation of photo-emf in a semiconductor surface layer

The *pn* junction of a photodiode separates the photocarriers generated by the incident radiation (Fig. 2.8*b*). This causes a decrease in the internal potential barrier (the effect is similar to that produced by the forward voltage on the normal diode). The Fermi levels on both sides of the *pn* junction no longer coincide, and one becomes higher than the other. It is of importance to point out that the radiant flux disturbs the equilibrium state in 'favor' of the drift component. This is indicative of the need to suppress the diffusion component in the photodiode to enable a more effective optical control of the photocurrent. This component is parasitic in the photodiode.

The drift flow of photocarriers produces a current called the photocurrent I_P. The holes charge the *p* region positively with respect to the *n* region, and the electrons charge the *n* region negatively with respect to the *p* region. The resultant potential difference, referred to as the photoelectromotive force E_P, lowers the internal potential barrier to ΔE (see Figs. 2.7 and 2.8*b*).

There are limits on the accumulation of photocarriers in the corresponding regions: concurrent with the storage of holes in the *p* region and electrons in the *n* region, there occurs a decrease in the potential barrier E_0 by the value of the induced photo-emf E_P. Obviously, E_P must be lower than E_0 because the separation of photocarriers is possible only when the potential barrier is present. A decrease in the height of the potential barrier with an attendant decrease in the strength of the electric field across the *pn* junction impairs the 'separation properties' of the junction.

Along with the main component of the photo-emf that arises from the separation of charge carriers by the *pn* junction field, other components may also be present. One of these is a Dember photo-emf. In the surface layer of an illuminated semiconductor there appear photocarriers—secondary holes and electrons (Fig. 2.9). The diffusion constant for electrons is higher than that for holes. The electrons diffusing from the site of carrier generation thus get ahead of the holes, leaving the holes behind, which form a positive charge on the surface. The potential difference between the surface layer and the semiconductor bulk is the Dember photo-emf. Besides, when the surface of a semiconductor has traps for carriers of one sign (only electron traps or only hole traps), the diffusion of carriers of opposite

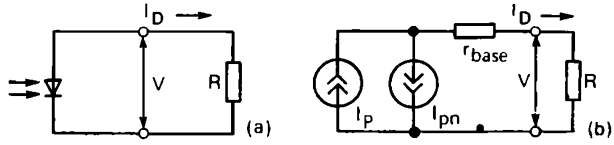


Fig. 2.10. Photodiode operating in the photovoltaic mode (*a*) circuit connection; (*b*) equivalent circuit

sign into the semiconductor bulk may give rise to a component of the photo-emf.

2.3.2. Modes of operation. For the photodiode to be more responsive to the incident radiation, the diffusion component must be kept at a minimum. To achieve the aim, the photodiode is made to operate either without any external bias at all (in the photovoltaic mode) or with a reverse bias (in the photoconductive mode).

In the photovoltaic (PV) mode of operation, the photodiode thus needs no external voltage source; the diode acts as a photocurrent generator (Fig. 2.10*a*). The expression for the current I_D through the photodiode D can be derived using its equivalent circuit (Fig. 2.10*b*):

$$I_D = V/R = I_P - I_{pn} = I_P(\Phi) - I_0[\exp(V/\varphi_T) - 1] \quad (2.10)$$

where I_P is the current of photocarriers (photocurrent); R is the load resistance; I_{pn} is the current through the *pn* junction; V is the voltage across the diode; I_0 is the thermal current through the *pn* junction; $\varphi_T = kT/e$ is the thermal potential; k is the Boltzmann constant; e is the electron charge; and $I_0 \exp(V/\varphi_T - 1)$ is the expression for the VI characteristic of the *pn* junction.

For an open-circuited output ($R = \infty$) the equivalent circuit gives $I_P = I_{pn}$. From expression (2.10) it is then easy to estimate the open-circuit voltage which is equal to the photo-emf:

$$V_{oc} = \varphi_T \ln(1 + I_P/I_0) \quad (2.11)$$

With the load short-circuited ($R = 0$), the voltage on the photodiode is $V = 0$ and the short-circuit current is $I_{sc} = I_P$, which is the flow of photocarriers generated in the diode.

In the photoconductive (PC) mode, the photodiode operates from a reverse bias source V_{SR} placed in series with the diode (Fig. 2.11). In this mode, the potential barrier gets higher, so the *pn* junction current I_{pn} will be governed by the thermal current I_0 which flows through the diode in the absence of illumination. The current through the illuminated photodiode is

$$I_D = I_P - I_{pn} = I_P + I_0 \approx I_P \quad (2.12)$$

2.3.3. Voltage-current characteristics. In the general case (for V of any polarity), the current of a photodiode is defined by expres-

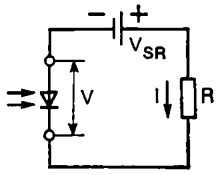


Fig. 2.11. Photodiode operating in the photoconductive mode

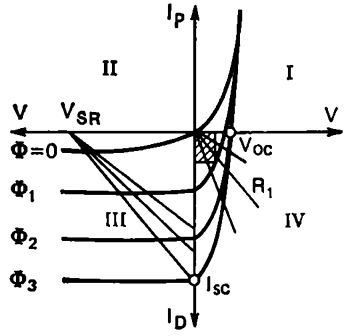


Fig. 2.12. *VI* characteristics of a photodiode

sion (2.10). This expression gives the current I_D as a function of voltage V across the diode with the radiant flux Φ as a parameter, i.e. it is the equation for the family of *VI* characteristics illustrated in Fig. 2.12.

The *VI* characteristics for a photodiode lie in quadrants I, III, and IV. Quadrant I is not an operating region for the diode: the forward bias on the *pn* junction gives rise to the diffusion component of current which totally swamps the photocurrent, $I_{pn} \gg I_P$. Optical control of the current through the diode is thus impossible.

Quadrant III is the photoconductive region of the diode operating with a reverse bias. It should be borne in mind that in the working range of reverse voltages, the photocurrent is practically independent of the reverse voltage and load resistance. The *VI* curve for the load resistor R becomes a straight line, whose equation has the form

$$V_{SR} - I_P R = V \tag{2.13}$$

where V_{SR} is the reverse voltage of the source; V is the reverse voltage across the diode; and I_P is the photocurrent (load current).

The same current I_P flows through both the diode and load resistor since they are in series. The estimate of this current can be made from the point of intersection of the *VI* curves for the diode operating in quadrant III and the load resistor. Thus the diode operating in the PC mode at the specified level of illumination is the source of photocurrent I_P with respect to the external circuit. The value of I_P is practically independent of the external circuit parameters such as V_{SR} and R.

Quadrant IV for the family of *VI* characteristics describes the operation of the photodiode in the PV mode. The points of intersection of the *VI* curves with the voltage axis correspond to the values of the photo-emf or open-circuit voltage ($R = \infty$) at different levels of illumination. In silicon photodiodes a typical value of the

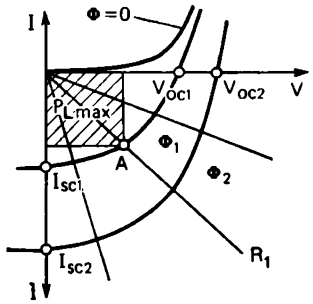


Fig. 2.13. *VI* characteristics of a photodiode in the photovoltaic mode of operation

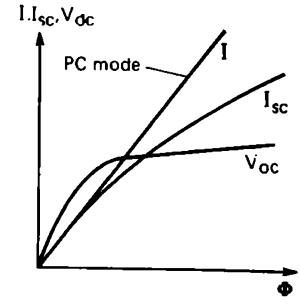


Fig. 2.14. I-Φ, I_{sc}-Φ, and V_{oc}-Φ curves for a photodiode in the PV and PC modes of operation

photo-emf is 0.50-0.55 V. The points of intersection of the *VI* curves with the current axis correspond to the values of the short-circuit current ($R = 0$). The intermediate values of the load resistance R_L lie on the load lines drawn from the origin at different angles depending on the values of R_L (Fig. 2.13). Given the value of current I, the *VI* characteristics can be used to select the optimum operating conditions for the diode as a photovoltaic device. The optimum operation here implies that the diode delivers the highest electric power P_{Lmax} to the load of the adequately selected resistance.

The line of load R_1 corresponds to the optimum operation for the flux Φ_1 (the area of the hatched rectangle with the apex at point A where the lines for Φ_1 and R_1 intersect will be the largest). For silicon devices operating into optimum loads, the voltage V lies between 0.35 and 0.40 V.

2.3.4. Energy characteristic. This characteristic relates the current I_P to the flux Φ incident on the surface of a photodiode. Since the radiant flux Φ falls on to the entire surface of the n region, the number of photons incident on this region in a unit time will be equal to $\Phi/h\nu$. The current I_P is obviously proportional to the number of photons absorbed within the semiconductor in a unit time.

$$I_P = q\eta\gamma_B \frac{\Phi}{h\nu} \tag{2.14}$$

where η is the internal quantum efficiency, i.e. the number of electron-hole pairs generated by one photon; q is the number of photons; and γ_B is the base transport factor which gives a fraction of charge carriers generated by the incident radiation that have escaped recombination on their way to the collector.

In the PC mode, the photodiode has an energy characteristic defined by expression (2.14), which is linear in the working range of the levels of illumination. This means that practically all photocar-

riers reach the *pn* junction and take part in the generation of photocurrent. In any case, the loss of minority carriers due to recombination is independent of the flux since the host material contains a small number of impurity particles which can act as recombination centers.

In the PV mode, the energy characteristic is a plot of either the short-circuit current I_{sc} or the photo-emf $E_P = V_P$ versus radiant flux. At large values of Φ, the I_{sc}-Φ or the V_{oc}-Φ relation deviates from the linear shape (Fig. 2.14). The I_{sc}-Φ relation becomes nonlinear primarily because the voltage drop across the bulk resistance r_b of the base grows with incident radiation. The photo-emf decreases because the height of the potential barrier lowers as the *n* region and *p* region store up excess charges of electrons and holes respectively; as a result, the *pn* junction field separates the electrons and holes less effectively and the growth of the photo-emf with increasing illumination slows down.

2.3.5. Spectral characteristic. Using the known relation $\nu = c_0/\lambda$, from expression (2.14) we can define the spectral responsivity S_P as a function of wavelength λ:

$$S_P = \frac{I_P}{\Phi} = \frac{q\eta\gamma_b\lambda}{hc_0} \tag{2.15}$$

From this expression it follows that the theoretical spectral response takes the shape of a straight line and passes through the origin (full line in Fig. 2.6). The practical spectral response is shown by a broken line in Fig. 2.6.

A decrease in the responsivity with shorter wavelengths is due to an enhanced energy absorption in the thin surface layer where the rate of recombination through the traps is much higher than in the bulk of the material. Hence, the short-wavelength cutoff of the spectral response for a photodiode depends on the base width and surface recombination rate. By decreasing these two parameters we can shift the short-wavelength cutoff into the region of shorter wavelengths. A decay of the response in the region of longer wavelengths appears at the absorption band edge, which is the long-wavelength cutoff of the spectral response of the material.

The peak of the practical response for a photodiode heavily depends on the absorption coefficient of the diode material. If the penetration depth χ_0 sharply decreases with a decrease in the wavelength of incident radiation, as is the case in germanium (see Fig. 2.4), the peak response becomes a function of the cutoff wavelength ($\lambda_{max} = 1.55$ μm for Ge) and does not practically depend on the base width. If χ_0 weakly depends on the wavelength, as is the case for silicon, the peak of the spectral response can shift with a change of both the base width and the rate of recombination. Thus in a sili-

con photodiode this peak can be shifted in the range from 0.6 to 1 μm.

2.3.6. Speed of response. The speed of a photodetector is a function of both the efficiency with which the *pn* junction field separates the carriers resulting from the absorption of radiation and the value of the depletion (junction) capacitance C_d. The electric field of the depletion layer separates the holes and electrons after they diffuse to the *pn* junction from the sites of generation. The transit time for carriers traversing the *pn* junction is

$$t_{tr} \approx \delta / v_{max}$$

where δ is the *pn* junction thickness; and v_{max} is the maximum carrier drift velocity.

In germanium and silicon, $v_{max} \approx 5 \times 10^4$ m s^{-1}. The *pn* junction thickness, which varies with the reverse voltage and impurity concentration in the base, is commonly less than 5 μm. The transit time is then smaller than 0.1 ns.

The time it takes for the depletion capacitance to recharge at a small value of the load resistance is a function of the time constant $C_d r_b$, where r_b is the base resistance of the diode. This time constant for common photodiodes is about 1 ns.

Calculations reveal that the speed of photodiodes mainly depends on the time the photocarriers take to diffuse to the *pn* junction. Keeping in mind this fact, consider the transients in the photoconductive mode of operation.

To determine how the load current varies with time, we need to solve a one-dimensional time-dependent equation for the diffusion of holes in the *n* region:

$$D_p \frac{\partial^2 (\Delta p)}{\partial x^2} - \frac{\Delta p}{\tau_p} = \frac{\partial (\Delta p)}{\partial t} \tag{2.16}$$

where $\Delta p = p - p_{n0}$ is the nonequilibrium hole concentration in the *n* base; τ_p is the lifetime of holes in the *n* base; and D_p is the hole diffusion constant.

Assume that at an instant $t = 0$ the excess hole concentration in all of the *n* region is $\Delta p = 0$. On illumination of the *n* region, nonequilibrium holes appear in the thin surface layer a distance w remote from the *pn* junction. The current produced by the flow of these holes is defined by expression (2.14). The boundary condition at $x = 0$ can then be written in the form

$$\frac{\partial (\Delta p)}{\partial x}\bigg|_{x=0} = -\frac{I_P}{qD_p} \tag{2.17a}$$

Since the photodiode operates with a reverse bias, the second boundary condition when $x = w$ assumes the form

$$\Delta p|_{x=w} = 0 \tag{2.17b}$$

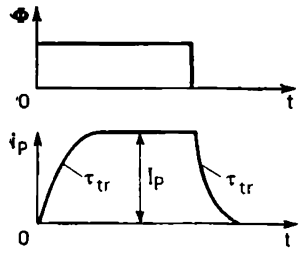


Fig. 2.15. Waveforms illustrative of the transient in a photodiode operating in the PC mode

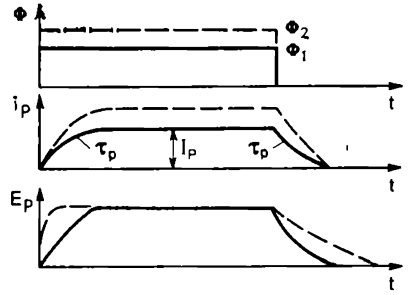


Fig. 2.16. Waveforms illustrative of the transient in a photodiode operating in the PV mode

The solution of the diffusion equation subject to the specified boundary conditions is written as

$$i_P = I_P [1 - \exp(-t/\tau_{tr})] \qquad (2.18)$$

where $\tau_{tr} = w^2/2D_p$ is the transit time for minority carriers in the n base.

With illumination turned off, the minority carriers are swept away across the pn junction and the current decays exponentially:

$$i_P = I_P \exp(-t/\tau_{tr}) \qquad (2.19)$$

The photocurrent waveform illustrative of the transient in the diode operating in the PC mode is shown in Fig. 2.15.

Let us look at the transients in the photovoltaic mode and determine how the photo-emf varies with time in the open-circuit conditions.

Let at an instant $t = 0$ a rectangular light pulse arrive at the photodiode. The pulse initiates the flow of minority carriers in the n region, the concentration of which increases with time. Since the hole current through the pn junction then begins to grow, the p region will continue to charge positively with respect to the n region. The rate of recombination of excess carriers will rise with an increase in their concentration. At a certain value of the carrier concentration, the recombination of carriers will counterbalance the carrier generation and the photo-emf will reach a steady value. After the light pulse has ended, the excess carriers recombine in the n region, their concentration decreases, and so does too the current I_P through the pn junction. The photo-emf also decreases, as is clear from expression (2.11). Thus in the PV mode the photo-emf and photocurrent vary with the rates of generation and recombination (storage and withdrawal) of excess carriers, the rate of each process being determined by the lifetime constant τ_p. As seen from

Fig. 2.16, the photocurrent rises in the manner defined as

$$i_P = I_P [1 - \exp(-t/\tau_p)] \tag{2.20}$$

and decays as

$$i_P = I_P \exp(-t/\tau_p) \tag{2.21}$$

The expressions for the rise and fall of the photo-emf can be found from equation (2.11) by substituting into it expressions (2.20) and (2.21):

$$E_P = \varphi_T \ln\left\{\frac{I_P}{I_0}[1 - \exp(-t/\tau_p)] + 1\right\} \tag{2.22a}$$

$$E_P = \varphi_T \ln\left[\frac{I_P}{I_0}\exp(-t/\tau_p) + 1\right] \tag{2.22b}$$

Figure 2.16 illustrates the waveforms plotted according to formulas (2.22) for various levels of illumination: it is seen that as the intensity of radiation grows, the rise time for the photo-emf decreases and the fall time increases.

For the PV mode, we thus have $\tau_{tr} = w_B^2/2D_p$, where w_B is the base width and D_p is the diffusion constant for minority carriers, which accounts for the time of diffusion of carriers from the sites of their generation to the *pn* junction. To reduce τ_{tr}, we need to reduce w_B. For silicon, $D_p = 0.01\ \mathrm{m^2\ s^{-1}}$ and $w_B = 10\ \mu\mathrm{m}$, i.e. $\tau_{tr} \approx 100$ ns.

A further decrease in w_B is impracticable, because the response then lowers and the response peak shifts into the short-wavelength region, which involves difficulties in matching the photodiode with the radiator. Besides, the effect of C_d becomes more appreciable. The use of more complex semiconductor structures made possible the fabrication of higher-speed photodiodes.

2.4. Photodiode Types

2.4.1. PIN photodiodes. The *pin* structure enables the diode to operate in a wider frequency band without reduction of its response (Fig. 2.17).

In this structure, the intrinsic (*i*) region lies between the two extreme regions of opposite conductivity types. A sufficiently high reverse voltage applied to the diode produces a strong and almost homogeneous field which expands into the entire *i* region. Since this medium region can be made rather wide, the *pin* structure offers the means for the fabrication of very fast detectors. Free carriers generated in the *i* region through the absorption of photons are rapidly swept off by the electric field toward the junction. The energy band model for a reverse biased *pin* diode is shown in Fig. 2.18. The *i* region has 10^6-10^7 times the resistivity of the doped *n*-type and *p*-type regions. This ensures a large width *w* of the space charge

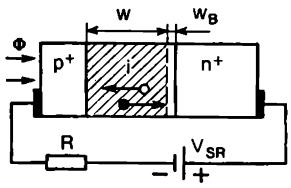


Fig. 2.17. Structure of a *pin* photodiode

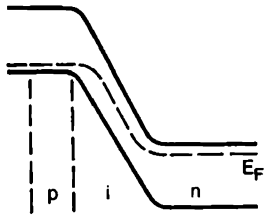


Fig. 2.18. Band diagram of a *pin* photodiode

region, enough to absorb the radiation directly within the confines of this region, where the external bias produces a strong electric field. The transit time is then equal to

$$t_{tr} = \frac{w}{v} = \frac{w}{\mu E} \tag{2.23}$$

where E is the electric field strength; μ is the carrier mobility; and $v = \mu E$ is the drift velocity of carriers in the electric field.

At a field strength of about 2×10^6 V m^{-1}, the carriers drift at maximum speeds of (6 to 8) $\times 10^4$ m s^{-1}. In this case, when $w = 10^{-2}$ cm, the transit time is 10^{-9} to 10^{-10} s. Large values of w and small values of w_b ensure a low product $C_d r_b$, from 10^{-11} to 10^{-12} s. The diode can cover a bandwidth of up to 10^9 Hz. This is a high-speed silicon diode.

Thus an increase in the speed of a *pin* diode is made possible by virtue of the drift of photocarriers in a strong *pn* junction field in contrast to the diffusion of carriers, as is the case in ordinary diodes.

The effectiveness of the drift process harnessed for the separation of light-generated carriers, as against the diffusion process specific to a *pn* photodiode, is seen from the comparison of the transit times for carriers crossing the base of width w:

$$\frac{t_{tr}(pin)}{t_{tr}(pn)} = \frac{w^2/\mu_p V_{SR}}{w^2/2D_p} \approx \frac{2kT/q}{V_{SR}} \approx \frac{0.05 \text{ V}}{V_{SR}} \tag{2.24}$$

because $D_p/\mu_p = kT/q$.

Consequently, even starting at reverse voltages V_{SR} of 0.1 or 0.2 V, *pin* photodiodes prove superior to *pn* diodes in speed.

So, *pin* photodiodes offer the following main advantages.

1. High current responsivity in combination with high speed (at a wavelength of about 0.9 μm, diodes exhibit S_P of about 0.7 A W^{-1}, which is near the theoretical limit).

2. High responsivity in the long-wavelength region by way of increase in the width of the intrinsic region.

3. Small values of the depletion capacitance.

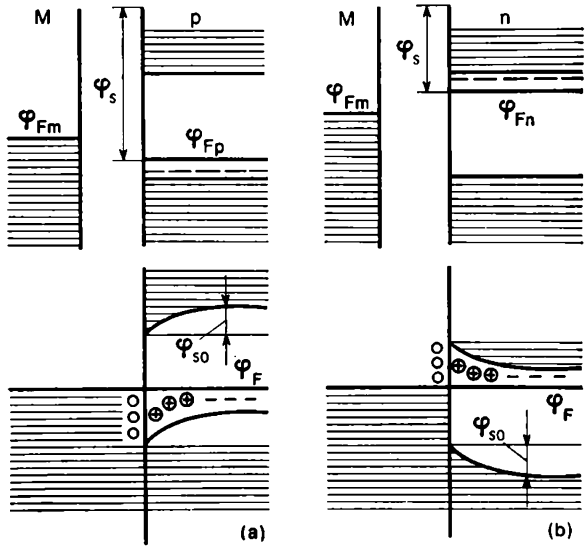


Fig. 2.19. Potential distribution diagrams for a metal-semiconductor junction (*a*) metal-*p*-type semiconductor junction; (*b*) metal-*n*-type semiconductor junction

4. Low operating voltages in the PC mode, which ensures electrical compatibility of *pin* diodes with integrated microcircuits.

As regards the disadvantages of *pin* diodes, mention should be made of high-purity requirements for the *i* region and poor fabrication compatibility with thin doped layers in integrated circuits.

2.4.2. Schottky barrier photodiodes. These are metal-semiconductor (MS) diodes which can provide a response of up to 10^{-10} s and even above. It is impossible to obtain a quality contact between the metal and semiconductor by merely joining them together. Practical metal-semiconductor contacts are now made by evaporation of metal in a vacuum to let it deposit on to the semiconductor.

The structure and properties of an M-S contact primarily depend on the mutual location of the Fermi levels in the metal and semiconductor. Figure 2.19 illustrates the potential distribution diagram for the case when the Fermi potential in a metal is higher than that in a *p*-type semiconductor, $\varphi_{Fm} > \varphi_{Fp}$. The difference in the Fermi potentials means that the probability of occupation of one and the same energy state is higher in the metal than in the conduction band of the semiconductor. For this reason the electron occupancy of the conduction band of the semiconductor is lower than that for the metal region having the same energy levels. In the ready metal-semiconductor structure a fraction of electrons will move from the metal to the *p*-type semiconductor. The excess electrons piled up in

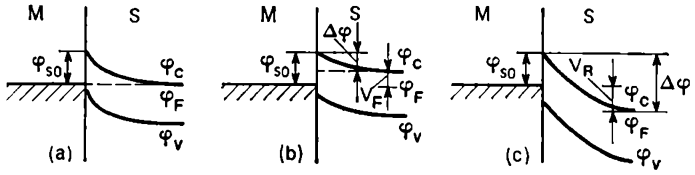


Fig. 2.20. Energy band diagrams for a Schottky photodiode (*a*) in equilibrium state; (*b*) under forward bias; (*c*) under reverse bias

the surface layer of the semiconductor cause an additional recombination process, so that the number of holes decreases and a negative space charge of ions (acceptors) appears at the M-S interface. The electric field so built up impedes the motion of electrons and secures a dynamic equilibrium in the junction region.

Figure 2.19*b* demonstrates the band models for the case when $\varphi_{Fm} < \varphi_{Fn}$ in the contact between the metal and *n*-type semiconductor. After formation of this structure, electrons leave the semiconductor for the metal. Positively charged ions (donors) and an electric field thus appear near the M-S interface.

Energy bands get bent near the semiconductor surface. The degree of band bending is defined by the surface potential φ_s, which is the potential difference between the bulk and the surface of the semiconductor. In the equilibrium state (when there are no external bias and illumination) the structure has an equilibrium surface potential φ_{s0}. The potential barrier in the interfacial (contact) layer is termed the Schottky barrier whose height φ_{s0} is analogous to the height of the internal potential barrier in the *pn* junction (Fig. 2.20*a*). The height of the Schottky barrier and the resistance of the interfacial layer will vary with the reversal of sign of the externally applied voltage.

With a forward voltage V across the contact (making the metal positive with respect to the *n*-type semiconductor), the potential barrier lowers, the interfacial layer gets enriched with majority carriers (electrons), and the M-S junction resistance drops below the equilibrium value (Fig. 2.20*b*). After changing the sign of external voltage to negative on M and to positive on S, the potential barrier across the junction gets higher (Fig. 2.20*c*). The interfacial layer then becomes more heavily depleted of majority carriers and the resistance tends to rise above the equilibrium value. The metal-semiconductor contact thus affords the rectifying action and can serve as a diode. The diodes using a Schottky barrier are known as Schottky diodes.

A feature specific to Schottky barrier diodes is that, unlike ordinary *pn* junction diodes, they do not inject minority carriers. The Schottky diode is said to 'operate with majority carriers' and it is free from retarding processes associated with the storage and clean-

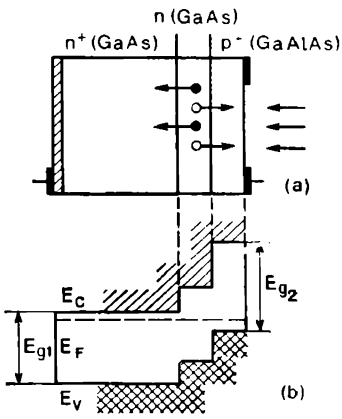


Fig. 2.21. Structure (*a*) and band diagram (*b*) of a heterojunction photodiode

out of minority carriers in the diode base. But this fact is of no importance for the practical Schottky photodiode since it does not operate with a forward bias.

As for the Schottky barrier photodiode, mention should be made of its two distinguishing features. First, the photodiode is capable of absorbing the photons of energy less than the bandgap energy on the metallic side of the contact. If the quantum energy is greater than the potential barrier, the excited electrons can move from metal to semiconductor through the potential barrier. The long-wavelength cutoff thus shifts into the region of longer wavelengths. Second, with an increase in the energy of quanta, the absorption region expands into the space charge layer having the field that separates electrons and holes. In a *pn* junction photodiode the photocurrent is generally equal to zero at a small light penetration depth. Consequently, the short-wavelength cutoff of the Schottky diode lies in the region of shorter wavelengths. The metal-semiconductor photodiode is thus responsive to a much wider range of wavelengths than the *pn* junction photodiode using the same semiconductor.

Other advantages of Schottky photodiodes are the following.

1. Small base resistance. As a result, the time constant $C_d r_b$ for the depletion capacitance is in the order of 10^{-12} s and the time lag is only dependent on the transit time the photocarriers take to traverse the space charge region, which lies between 10^{-10} and 10^{-11} s.
2. High speed combined with high responsivity, $S_P = 0.5$ A W^{-1}.
3. Ease of fabrication from a great variety of semiconductors including those which are impracticable for the *pn* junction diode and, as a result, the possibility of control of the Schottky barrier height.

2.4.3. Heterojunction photodiodes. These photodiodes belong to one of the most promising types of photodetectors. The heterostructure opens the way for fabrication of photodiodes whose efficiency would come close to 100%. The basic properties of heterostructures (heterojunctions) were dealt with in Sec. 1.3. Here we illustrate the features of heterodiodes by an example of the GaAs-GaAlAs heterostructure. The schematic representation and the band diagram of a heterojunction photodiode appear in Fig. 2.21. The GaAlAs layer plays the role of a window transmitting the radiant flux to be absorb-

ed in the intermediate n region (GaAs). The difference in the energy of bandgaps on both sides of the heterojunction is about 0.4 eV. Light-generated holes freely travel from the n to the p^+ region. The width of the active n region is so chosen as to enable the material to absorb the entire radiation. A high degree of purity of the active region and a low density of surface states at the boundaries between the layers ensure a small loss of photocarriers due to recombination.

Thus, along with an exceptionally high efficiency, the heterojunction photodiode offers the advantages of the above described structures, namely, a high speed due to a high responsivity and low operating voltages.

The heterostructure enables the photodiode to operate in practically any region of the optical spectrum, provided the pairs of semiconductors are properly chosen. The thing is that the choice of the working wavelength here only depends on the difference between the bandgaps and is independent of the penetration depth χ_0. Since there are good possibilities for selecting an adequate base material, these photodiodes can generate a photo-emf in the range of 0.8-1.1 V, which is 2 or even 3 times as high as that of silicon photodiodes. What puts heterojunction photodiodes at a disadvantage, just like any heterostructure in general, is that they call for a complex manufacturing process.

2.4.4. Avalanche photodiodes. One of the ways for the creation of high-speed photodetectors is to harness for use the phenomenon of avalanche breakdown. If the field in the active region of a photodiode is strong and the energy acquired by the photocarriers (electrons and holes) in this field exceeds the energy needed to generate electron-hole pairs, the avalanche multiplication of carriers sets in. The multiplication process begins with the generation of carriers due to absorption of light quanta.

Amplification of the primary photocurrent in an avalanche photodiode is defined by the avalanche multiplication factor

$$M = I_P/I_{P0} \tag{2.25}$$

where I_P is the output photocurrent resulting from avalanche multiplication; and I_{P0} is the current in the absence of multiplication. This factor is thus the photocurrent gain.

The multiplication factor is known to be a function of the voltage across the junction:

$$M = \frac{1}{1-(V/V_{br})^m} \tag{2.26}$$

where V_{br} is the breakdown voltage; V is the voltage across the pn junction; m is a figure of merit equal to 1.5-2.0 for p-type silicon and 3.4-4.0 for n-type silicon. The VI characteristic for an avalanche

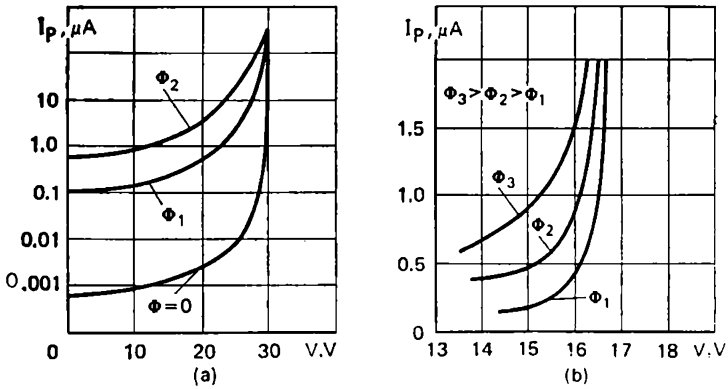


Fig. 2.22. *VI* characteristics of avalanche photodiodes

photodiode can then be described by the expression of the form

$$I_P = \frac{I_{P0}}{1-(V/V_{br})^m} \tag{2.27}$$

Figure 2.22 displays the *VI* curves typical for an avalanche photodiode.

Avalanche photodiodes are applicable for detection of weak optical signals. Wider uses of these diodes involve appreciable difficulties. This is primarily because the photocurrent gain M of the diode operating in the near-breakdown region of its characteristic heavily depends on voltage. Indeed, from (2.26) we can obtain the expression for the calculation of a relative change in the photocurrent gain with a change in the voltage on the junction:

$$\frac{dM}{M} = \frac{mM}{V}\, dV \left(\frac{V}{V_{br}}\right)^m \approx mM \frac{dV}{V} \tag{2.28}$$

The photodiode can obviously operate in the normal fashion only when the output current I_P and, hence, the gain M are sufficiently stable. Assume the relative change in M must be not higher than 10%, i.e. $dM/M = 0.1$. For the silicon photodiode with $m = 3.5$ and $M = 300$, we get $dV/V = 0.0001 = 0.01\%$. As is clear, the stability of the supply voltage must be very high for the photodiode to operate reliably. Moreover, since the typical values of V_{br} lie between 30 and 100 V, the photodiode dissipates a large amount of energy, $V_{br}I_P$. The breakdown builds up in separate 'microplasmas' rather than at once in the entire *pn* junction area. This leads to an additional instability of M and increases the level of noise. The above disadvantages along with the spread of parameters in individual specimens limit the area of application of avalanche photodiodes.

2.5. Current-Amplifying Photodetectors

These devices not only convert radiant energy to photocurrent but also amplify this photocurrent. The main varieties include a phototransistor, composite phototransistor, and photothyristor. Generally speaking, the avalanche photodiode described in Sec. 2.4 also belongs to this group of devices.

Figure 2.23 presents the typical structures of these photodetectors. The radiant flux falling on the active region of the structure through the input window initiates the flow of photocarriers separated by the *pn* junction. The photocurrent so produced is amplified by the normal transistor action.

2.5.1. Phototransistor. Fig. 2.23*a* illustrates a phototransistor generally connected into an external circuit in a common-emitter configuration. The base may have no external terminal, i.e. it may be left floating, so the transistor operates as a two-terminal device. Modern phototransistors often have an external base terminal.

Consider the operation of a phototransistor illustrated in Fig. 2.24. The absorption of radiation causes the generation of carriers in the base, which are separated by the collector junction. The holes cross the junction and move into the *p* region, while the electrons remain in the base. The field produced by the space charge of electrons cannot reduce the charge in the base at the expense of the base current ($I_B = 0$). The space charge field thus lowers the potential barrier of the emitter junction, causing an additional injection of holes into the base. The photocurrent plays here the role of the base current. The output characteristics of the phototransistor are consequently similar to those of a bipolar transistor (Fig. 2.25). In comparison with a common photodiode, the phototransistor affords current am-

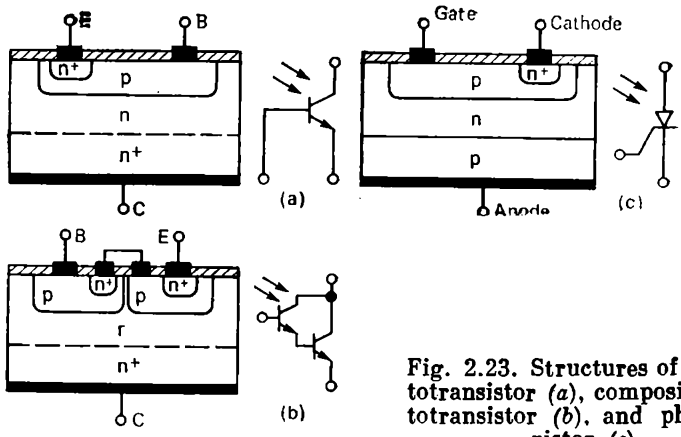


Fig. 2.23. Structures of a phototransistor (*a*), composite phototransistor (*b*), and photothyristor (*c*)

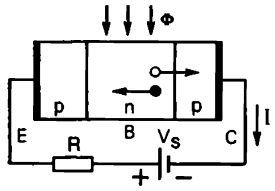


Fig. 2.24. Principle of a phototransistor

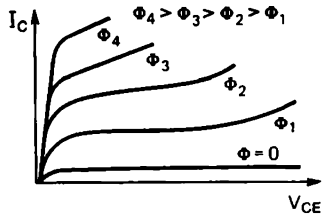


Fig. 2.25. Output characteristics of a phototransistor

plification by β times and exhibits the total responsivity equal to

$$S_P = S_D\beta \tag{2.29}$$

where S_D is the current responsivity of a photodiode formed by the emitter junction of the transistor; and β is the common-emitter current gain.

An enhanced responsivity is the main advantage of the phototransistor over the photodiode, although this advantage is brought about at a sacrifice in the temperature stability of the device since it operates with a dc bias on the base.

Phototransistors also show a decrease in the threshold response due to an appreciable increase in the dark current

$$I_d = I_{C0}\,(1 + \beta) \tag{2.30}$$

In particular, the following circumstance limits the area of application of phototransistors and impedes further improvements of their parameters: the requirements for a high transfer ratio and a low switching time call for a decrease in the base width w_b, which generally tends to reduce the device photosensitivity. The tradeoff between the opposing requirements necessitates lowering the speed of phototransistors, so that t_{sw} usually lies between 10^{-5} and 10^{-6} s.

It is possible to raise the speed of integrated photodetectors by combining a photodiode and transistor to form a single functional unit (Fig. 2.26). The separate optimization of the two structures enables the fabrication of a high-speed photodiode and high-frequency transistor on the same chip, as illustrated in Fig. 2.26*a*. This device makes a high-speed high-gain phototransistor.

2.5.2. Composite phototransistor. This is a photodarlington (see Fig. 2.23*b*) that affords a still higher response because its total current gain β greatly exceeds that of either of the two components. The relation between the currents in the photodarlington has the form (Fig. 2.27)

$$\begin{aligned} I_C &= I_{C1} + I_{C2} = \beta_1 I_{B1} + \beta_2 I_{B2} \\ &= \beta_1 I_{B1} + \beta_2\,(1 + \beta_1)\,I_{B1} \end{aligned} \tag{2.31}$$

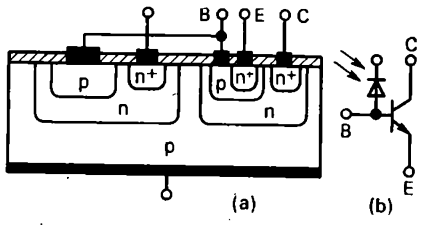


Fig. 2.26. Photodiode-transistor structure

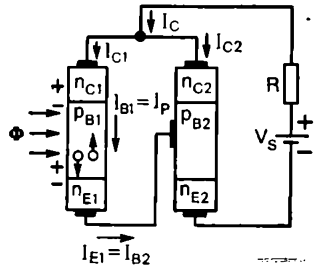


Fig. 2.27. Currents in a composite phototransistor

The current gain is equal to the ratio I_C/I_{B1}. Consequently

$$\beta = I_C/I_{B1} = \beta_1 + \beta_2 + \beta_1\beta_2 \approx \beta_1\beta_2 \qquad (2.32)$$

As a result, composite phototransistors commonly have over 1 000 times the responsivity of photodiodes.

2.5.3. Photothyristor. This is a four-layer *pnpn* semiconductor switch which can have two, three, or four terminals. The two-terminal device is called a switching diode, or dynistor. The three-terminal device is known as a light-activated silicon controlled switch (SCS). Figure 2.28 illustrates the structure of a diode photothyristor operating with a forward bias so that the anode is made positive with respect to the cathode. In the static conditions when direct current flows through the device, we can apply the principle of current continuity and express the current passing through the central junction J_2 in the form

$$I_{J2} = (I_{C0} + I_{P2}) + (I + I_{P1})\,\alpha_1 + (I + I_{P3})\,\alpha_2 \qquad (2.33)$$

whence,

$$I\,(1 - \alpha_1 - \alpha_2) = I_{C0} + I_{P1}\alpha_1 + I_{P2} + I_{P3}\alpha_2 \qquad (2.34)$$

where I_{P1}, I_{P2}, and I_{P3} are the photocurrents resulting from the separation of light-generated electron-hole pairs by respective *pn* junctions; α_1 and α_2 are the emitter current gains (common-base current transfer ratios) for the $p_1n_1p_2$ and the $n_2p_2n_1$ transistor respectively.

With no illumination of the diode photothyristor, i.e. when $I_{P1} = I_{P2} = I_{P3} = 0$, we obtain the expression for the *VI* characteristic of the device, i.e. the expression which specifies its dark-current characteristics. In the illuminated device, the current I flowing through the structure will result from the cumulative action of photocurrents I_P through the junctions and the intrinsic current through the collector (central) junction, I_{C0}. It can be said that the quantity $I_{P1}\alpha_1 + I_{P2} + I_{P3}\alpha_2$, which varies with the level of illumination, plays the role of a control current in the ordinary thyr-

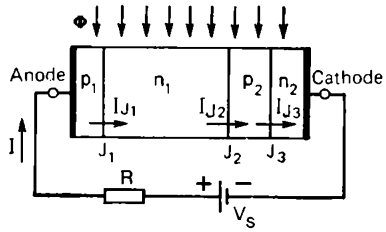


Fig. 2.28. Photothyristor structure

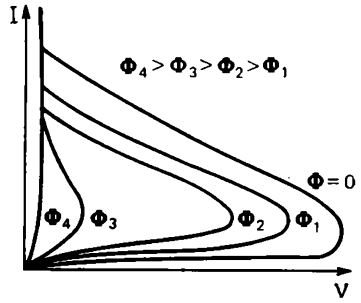


Fig. 2.29. *VI* characteristics of a photothyristor

istor; in other words, the switching voltage of the photothyristor changes under illumination.

Figure 2.29 shows the set of *VI* characteristics for a photothyristor illuminated by monochromatic light of different levels.

2.5.4. Comparison of characteristics. In conclusion, compare the parameters of photodetectors from the viewpoint of their applicability. Phototransistors excel other types of photodetector in a number of properties. First of all, they can be made in a variety of circuit configurations capable of performing a wide range of functions. An example of the functional flexibility of these devices is that they allow for the output current control both optically (by the radiant flux) and electrically (through the base circuit). The output circuit of a phototransistor can operate in linear and switch modes. Phototransistors have much higher output currents than, say, photodiodes and thus can function as switches in a wide range of electronic circuits. Composite phototransistors exhibit the highest responsivity and the highest values of switching currents and voltages but have a longer switching time, up to 100 μs. Photodiode-transistor structures (see Fig. 2.26) excel other photodetectors in switching time (about 1 μs) and current transfer ratio K around 10 per cent. The phototransistor proper has the values of responsivity and speed that are intermediate between these values for the photodiode/transistor and photodarlington.

Photothyristors are exceptionally suitable for use as high-power switches, in which respect they practically have no rivals.

2.6. Photoresistors

Photoresistors principally differ from other photodetectors in physical and design-fabrication features and also in the range of parameters and fields of application. This explains why all aspects relating to photoresistors are dealt with in the separate section.

2.6.1. Photoconduction. The photoresistor relies on the photoconductive effect which shows up as a change in the conductivity of a semiconductor when exposed to various amounts of radiation. Consider the photoconductive effect in more detail.

Nonequilibrium electrons and holes generated by photons may acquire energy at the instant of ionization that is much in excess of the mean energy of equilibrium carriers (about one kT). But as they interact with phonons and the crystal lattice, the photocarriers rapidly heat up to the temperature of the lattice. The energy of photocarriers becomes equal to the mean thermal energy of equilibrium carriers. This process takes about 10^{-10} s, i.e. the time that is a few orders of magnitude lower than the lifetime of carriers. Therefore, the energy distribution of nonequilibrium carriers can be assumed to correspond to that of equilibrium carriers and, hence, the mobilities of these carriers remain the same. The generation of carriers by the action of light thus leads to a change in the conductivity σ of the semiconductor:

$$\sigma_P = \sigma - \sigma_0 = q(\mu_n \Delta n + \mu_p \Delta p) \tag{2.35}$$

It is obvious that Δn and Δp depend on the time and level of illumination of the semiconductor.

Denote the rates of generation of electrons and holes by the incident radiation as $\Delta n'$ and $\Delta p'$ respectively. They must certainly be proportional to the light energy absorbed in a unit volume of the semiconductor per unit time. If the flux Φ of monochromatic radiation falls on the layer of thickness dx and the light penetration depth is χ_0, the light energy absorbed in a unit volume per unit time, as is evident from (2.3), will be equal to

$$-\frac{d\Phi}{dx} = \frac{\Phi}{\chi_0} \tag{2.36}$$

The carrier generation rate $\Delta n' = \Delta p'$ is thus proportional to Φ/χ_0. For the self-absorption region,

$$\Delta n' = \Delta p' = \eta_e \Phi/\chi_0 \tag{2.37}$$

where η_e is the internal quantum efficiency which accounts for the number of electron-hole pairs or, in the case of impurity photoconduction, the number of carriers generated by one absorbed photon; η_e is commonly much smaller than 1.

In a semiconductor illuminated by a constant flux, we have $\Delta p = \Delta n = \text{constant}$.

Estimate $\Delta n\,(t)$ and $\Delta p\,(t)$ at the constant value of Φ in the entire bulk of the specimen, which causes a uniform generation of charge carriers.

At the time $t = 0$, i.e. immediately at the beginning of illumination, the rate of recombination starts to rise with the concentration

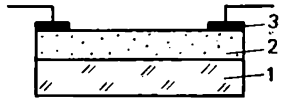


Fig. 2.30. Photoresistor structure

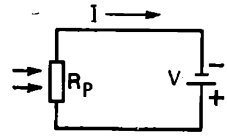


Fig. 2.31. Photoresistor connected in an electric circuit

of nonequilibrium carriers of charge. Since the rate of generation of nonequilibrium carriers remains constant at the constant flux, the recombination rate grows fast and reaches the carrier generation rate. This results in the static state for the nonequilibrium carrier concentration.

A change in the nonequilibrium carrier concentration per unit time is thus the difference between the rates of generation and recombination of carriers:

$$\frac{d(\Delta n)}{dt} = \eta_e \frac{\Phi}{\chi_0} - \frac{\Delta n}{\tau} \tag{2.38}$$

Setting $\Delta n = \Delta p \ll p_0 \gg n_0$, we may assume that τ is constant, so that

$$\Delta n(t) = \tau \eta_e (\Phi/\chi_0)[1 - \exp(-t/\tau)] \tag{2.39}$$

The steady-state value of the nonequilibrium concentration when $t \to \infty$ is equal to

$$\Delta n_0 = \tau \eta_e (\Phi/\chi_0) \tag{2.40a}$$

With the light instantly switched off, the nonequilibrium concentration starts decaying according to the expression

$$\Delta n(t) = \Delta n_0 \exp(-t/\tau) = \tau \eta_e (\Phi/\chi_0) \exp(-t/\tau) \tag{2.40b}$$

The nonequilibrium carrier concentration thus rises and drops off at the instantaneous turn-on and turn-off of the light with the lifetime constant τ.

Besides, we may conclude that where the recombination proceeds linearly, in which case the lifetime of nonequilibrium carriers is independent of the flux of radiation, the nonequilibrium steady-state concentration varies linearly with the flux in compliance with expression (2.37).

2.6.2. Parameters and characteristics. The basic element of a photoresistor, whose structure is shown in Fig. 2.30, is a thin semiconductor layer *2* deposited on a dielectric substrate *1*. Metallic terminals *3* provide connection with the external circuit.

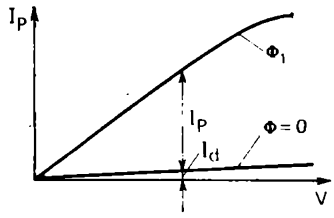


Fig. 2.32. *VI* characteristic of a photoresistor

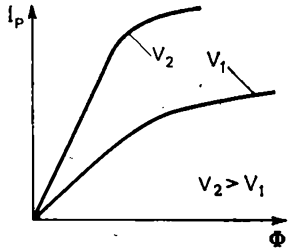


Fig. 2.33. I_P through a photoresistor versus flux Φ

With the photoresistor connected in the circuit as illustrated in Fig. 2.31, only the dark current flows in the absence of illumination:

$$I_d = b\sigma_0 V \tag{2.41}$$

where b is a constant accounting for the geometrical dimensions of the photosensitive layer; and $\sigma_0 = q(\mu_p p_0 + \mu_n n_0)$ is the dark conductivity.

The radiation falling on the photoresistor decreases the resistance, thus producing the photocurrent

$$I_P = b(\sigma_0 + \sigma_P) V \tag{2.42}$$

where σ_P is the light-induced conductivity (photoconductivity).

The *VI* characteristic is thus linear in form when the voltage across the photoresistor is not too high (Fig. 2.32). Figure 2.33 illustrates one more important characteristic of photoresistors. This is the energy characteristic which defines the photocurrent as a function of the incident radiation, $I_P = f(\Phi)$. At small values of the flux the photocurrent varies linearly with Φ, $I_P \sim \Phi$, but at high values of Φ the relation takes the form $I_P \sim \sqrt{\Phi}$.

Photoresistors find limited uses, although they are simple in structure and small in size and mass. This is because they have a large switching time, 10^{-2} to 10^{-3} s, poor linearity of the I_P-Φ characteristic, and parameters that are heavily dependent on temperature.

2.7. Photodetectors Using the Charge-Coupled Concept

In the field of imaging, it was not until quite recently that the solid-state devices began to edge out the cathode-ray tube (CRT). The breakthrough came with the advent of photosensitive charge-coupled devices (CCDs). Solid-state imagers (video signal generators) based on optical CCDs are successfully taking over from classical visible image pickup tubes of the vidicon type, bringing forth a numb-

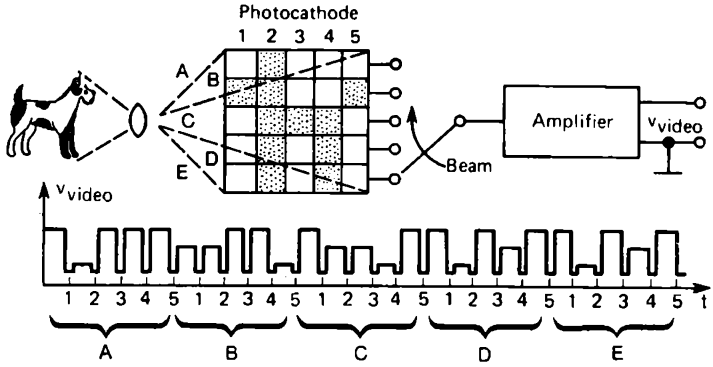


Fig. 2.34. Conversion of optical signals to video signals in a cathode-ray tube

er of advantages such as a decreased size and mass, improved reliability, and lower power per circuit. Besides, as compared to CRTs, CCD imagers embody such a principally new feature as a rigid raster and require low supply voltages. CCD array imagers are now available, which satisfy the requirements of TV specifications. The development of CCD-based display units continues with much success. In 1980 CCD color TV cameras were used for shooting at the ocean bottom. In 1981 the first Soviet TV cameras based on multielement CCDs for video tape recorders appeared on the market.

To get a better understanding of the operation of an optical CCD, consider first the principle of conversion of an optical signal to an electric signal (video signal) in a CRT (Fig. 2.34). The objective lens projects an image on the light-sensitive photocathode formed by a large number of microphotodetectors. The electron beam traveling inside the CRT rapidly sweeps over these elementary photodetectors and collects the electric charges they have stored. The more intensive the illumination of a particular photodetector, i.e. the brighter the image spot that has fallen on the photosite, the larger the charge it will store and thus the stronger the current that will appear at the CRT output at the instant of readout. On scanning across the entire photocathode, point by point and line by line, the electron beam produces a video signal which carries an electrical replica of the visible scene. Thus the process of conversion of an optical signal to a video signal consists of two stages: image recording on to the photocathode and image sensing (readout) by the electron beam.

The aim of converting from the CRT to a solid-state image pickup device was to implement the functions of image recording and readout in a semiconductor crystal. The problem was brought under control with the invention of optical CCDs.

2.7.1. MOS capacitor. In its simplest form, the image sensing CCD is an array of interacting metal-oxide-semiconductor (MOS)

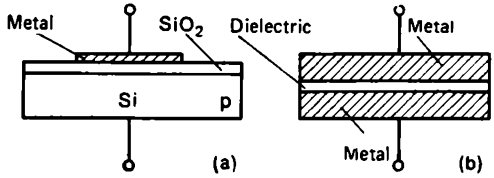


Fig. 2.35. Structures of a MOS capacitor (*a*) and ordinary capacitor (*b*)

structures arranged on a common substrate. As illustrated in Fig. 2.35*a*, the MOS structure consists of a semiconductor (commonly silicon), dielectric (most often silicon dioxide), and a metal plate. From the comparison of the MOS structure with a common capacitor (Fig. 2.35*b*) it is seen that the former is a capacitor too, only here the semiconductor layer replaces the lower metal plate. Because of the specific properties of a semiconductor, the MOS capacitor differs in principle from an ordinary capacitor. Its features follow from the field effect in the semiconductor, which manifests itself as a carrier concentration change in the semiconductor surface layer in response to an electric field. A dc voltage impressed across the MOS capacitor will induce equivalent charges on its plates. But unlike the charge on the metal plate, the charge in the semiconductor does not accumulate at its surface, but spreads out for a certain distance into the semiconductor bulk. Assume the semiconductor is silicon of the *p* type. If we apply a negative voltage to the metal plate, called the gate, the holes will collect in the surface layer of the semiconductor which is said to be in the enhancement state. The semiconductor can then be regarded as a metal plate, so the per-unit area capacitance (capacitance per unit area of the gate) can be taken equal to

$$C_0 = \frac{\varepsilon_d \varepsilon_0}{d} \tag{2.43}$$

where ε_d is the relative permittivity of the dielectric; and d is the dielectric layer thickness.

If we apply a small positive voltage to the gate, the semiconductor will collect a negative charge because the positive voltage causes the holes to repel from the semiconductor surface. Since each elemental volume of the semiconductor contains a strictly definite number of acceptor impurity atoms, there appears a surface layer of thickness l_0 that is free of mobile holes. The layer carries a negative charge whose density is

$$\rho = -qN_a \tag{2.44}$$

where N_a is the acceptor concentration. This layer is known as a depletion region and the MOS capacitor in which this layer appears is said to be in a depletion state, or mode. This is the main operation mode for optical CCDs.

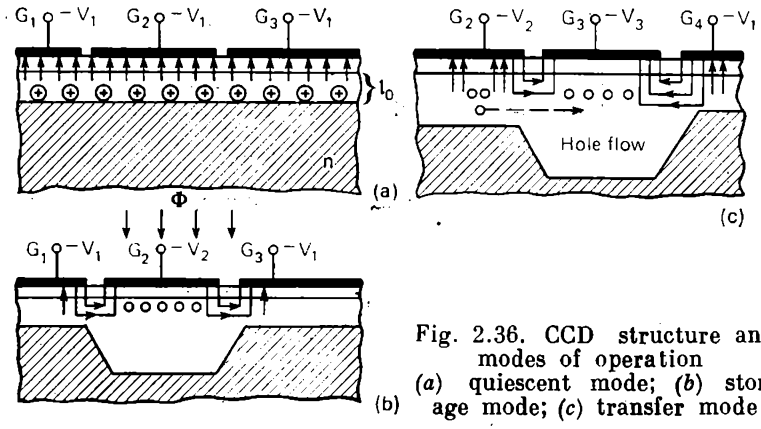


Fig. 2.36. CCD structure and modes of operation (*a*) quiescent mode; (*b*) storage mode; (*c*) transfer mode

The potential distribution inside the depletion region can be found from the solution of Poisson's equation:

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon_{Si}\varepsilon_0} = \frac{qN_a}{\varepsilon_{Si}\varepsilon_0} \tag{2.45}$$

where ε_{Si} is the relative permittivity of silicon.

Integrating equation (2.45) subject to boundary conditions $\varphi = 0$ and $d\varphi/dx = 0$ when $x = l_0$ yields

$$\varphi(x) = \frac{qN_a}{2\varepsilon_{Si}\varepsilon_0}(x - l_0)^2; \quad 0 \leqslant x \leqslant l_0 \tag{2.46}$$

where l_0 is the depletion region thickness (the point $x = 0$ lies at the interface between semiconductor and dielectric).

The potential at the interface ($x = 0$) is the surface potential

$$\varphi_s = \frac{qN_a}{2\varepsilon_{Si}\varepsilon_0} l_0^2 \tag{2.46a}$$

whence the depletion region thickness (depth) is

$$l_0 = \sqrt{\frac{2\varepsilon_{Si}\varepsilon_0\varphi_s}{qN_a}} \tag{2.46b}$$

2.7.2. Charge packet and CCD gate operation modes. A CCD consists of a large number of interacting MOS capacitors arranged adjacent to each other. Assume the CCD semiconductor is of the *n* type and all gates are fed with negative voltages V_1 of the same magnitude (Fig. 2.36*a*). The depletion region of depth l_0 then appears under each gate. This region contains a positive space charge of donor impurity ions. If the voltage on the gates is the same, the electric field obviously penetrates the semiconductor to the same depth along the entire structure.

If the voltage on the gate G_2 (Fig. 2.36*b*) is more negative than that on the adjacent gates G_1 and G_3, i.e. $|V_2| > |V_1|$, the depletion region depth will become greater under G_2 than under G_1 and G_3. Let the semiconductor under G_2 be momentarily illuminated by a flux Φ. This flux generates holes and electrons in the semiconductor, and the electric field separates these photocarriers: electrons move into the bulk of the semiconductor and holes accumulate at its surface, forming what is called a hole charge packet. The charge packet grows in magnitude with the amount of radiant energy and can thus serve as a carrier of information on an image.

When the voltage V_2 on G_2 is more negative than V_1 on G_1, the charge packet is unable to leave the region under G_2 since the electric field acting at its boundaries retards the holes. The charge can be stored under G_2 for a relatively long time, in which case the gate G_2 is said to operate in the storage mode at the storage voltage V_2.

Let the voltage V_3 applied to the gate G_3 be more negative than V_2 on G_2 (Fig. 2.36*c*). An accelerating field will then appear at the boundary between G_2 and G_3. This field speeds up holes, so the charge packet moves from the gate G_2 toward the gate G_3, which is said to operate in the readout (transfer) mode at the readout (transfer) voltage V_3.

The total positive charge under a gate is a function of the gate voltage, $Q = CV$. For this reason, the buildup of the hole packet entails a decrease in the positive charge of depleted donors. With the hole packet and the positive charge of depleted donors being balanced out completely, the retarding fields at the boundaries between the gates disappear and the hole packet spreads along the entire surface. A maximum permissible charge of the hole packet has the form

$$Q_{\max} = (V_2 - V_1)\, C_0\, (ZL) \tag{2.47}$$

where C_0 is the per-unit area capacitance defined by (2.43); Z and L are the gate width and length respectively. The value of $Q_{\max}$ determines the control ability of a CCD and thus its resolution ability as a photodetector.

During storage of the hole packet under a gate, additional holes tend to appear as the result of thermal generation of carriers. The charges under all gates may then gradually level off, so the notion of a charge packet that underlies the principle of CCD operation becomes meaningless. Consequently, the storage time has an upper limit and the CCD operates in the unsteady, dynamic condition that does not allow for a long-term storage of information.

In the course of charge packet transfer, the total charge suffers losses due to two causes. First, the speed of charge flow from gate to gate is not uniform and a certain time is required to transfer all of the charge without loss. Second, a fraction of carriers drop into sur-

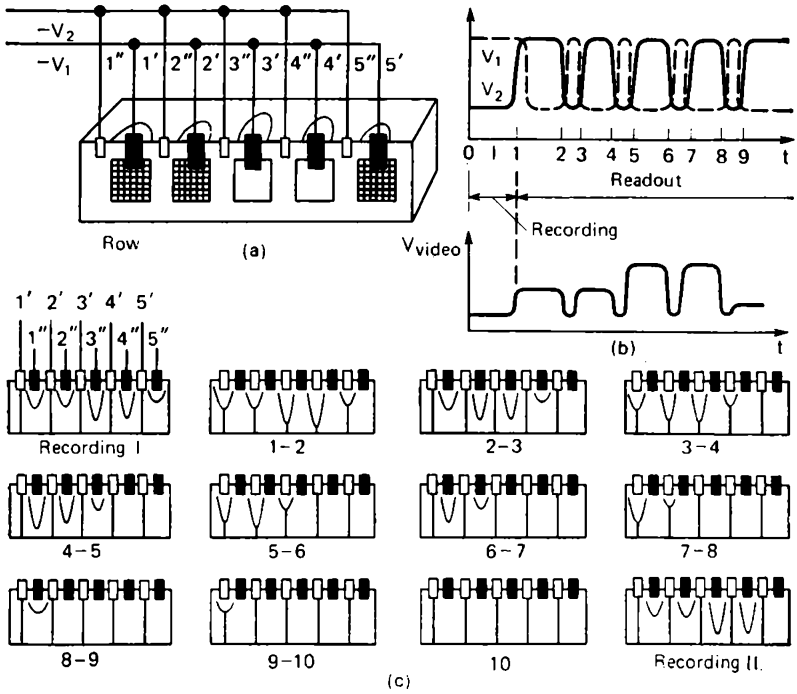


Fig. 2.37. Conversion of optical signals to video signals in a CCD row (*a*) CCD row storing electrical replica of scene; (*b*) waveforms of changes in control voltages and video signal; (*c*) charge packet transfer steps

face traps and cannot escape them during transfer. Thus a loss-free charge packet transfer requires a certain minimum time. This time decreases with a decrease in the spacing between the gates and an increase in the carrier mobility and readout voltage.

2.7.3. Data recording and readout. Consider the principle of image sensing in the CCD row comprising five MOS capacitors (Fig. 2.37*a*). The light passing through a transparent dielectric window falls on a semiconductor and generates there photocarriers (holes and electrons). On applying a negative voltage V_1 to electrodes *1′* through *5′*, the light-induced holes will be drawn to these electrodes and the electrons will be pulled into the crystal bulk. Let us now apply a voltage V_2 to electrodes *1″* through *5″* that is more negative than V_1 (at instant *1* in Fig. 2.37*b*). This voltage will produce an electric field between each pair of electrodes, *1′-1″*, *2′-2″*, *3′-3″*, etc., the field being directed from a more positive electrode (with one prime) to a more negative electrode (with a double prime), i.e. from right to left. The field causes the holes to shift to the left from electrodes *1* through *5* to electrodes *1″* through *5″*, i.e. from *5′* to *5″*, from *4′*

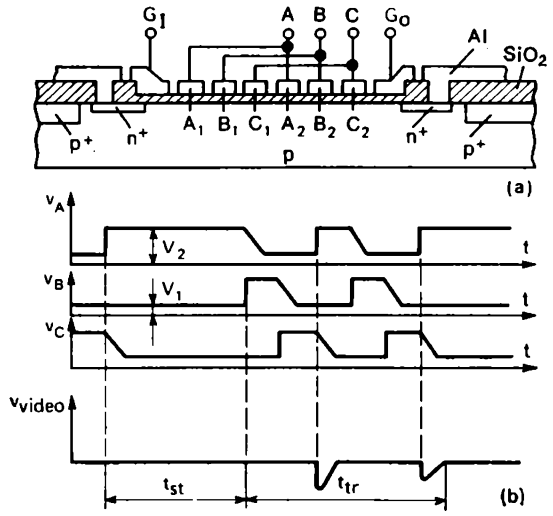


Fig. 2.38. Optical CCD using a three-phase control circuit
(*a*) CCD structure; (*b*) waveforms of control voltages and video signal

to *4″*, etc. as illustrated in Fig. 2.37*c*. The charge stored under the electrode *1* moves to *1″* and then to the input of a video amplifier, thereby producing the first component of the video signal (see Fig. 2.37*b*) which carries information on a definite spot of the image. By varying in each successive step the values of V_1 and V_2, as illustrated in Fig. 2.37*b*, it becomes possible to shift all holes to the left and obtain the entire video signal of the row. This brings to an end one cycle of recording and readout, and then the next cycle follows suit. The speed of recording and readout must obviously be much higher than the rate of change of the image. It should be pointed out that the CCD may comprise a huge number of MOS capacitors, but the readout process is run by the same pair of voltage, V_1 and V_2.

The described principle of readout is actually much more complex: with a change of voltages, the charge will not necessarily move leftward in the direction of the readout electrode *1″*; it is equally probable that the charges will move rightward because the voltages on *1″* and *2″* are the same. To ensure a unidirectional motion of charges, the CCD should incorporate a third group of electrodes, i.e. it should have a three-phase control circuit.

Figure 2.38 shows a three-phase *n*-channel CCD composed of six MOS capacitors connected to phase lines *A*, *B*, and *C* providing supply voltages. The waveforms illustrative of changes in the supply voltages appear in Fig. 2.38*b*. In the storage (recording) interval t_{st}, the line *A* is fed with voltage V_2, the lines *B* and *C* being supplied with voltage V_1 ($V_2 > V_1$). All MOS capacitors are driven into

deep depletion in this time interval, and the surface potentials differ in magnitude depending on the voltage applied to the electrodes. The surface potential under the electrodes connected to the line A will be higher than that under the electrodes of B and C. If the spacing between the electrodes is small, the gradation from the high surface potential under electrodes A_1, A_2 to the low surface potential under electrodes B_1, B_2, C_1, C_2 will be smooth, without potential barriers.

There appear potential wells under electrodes A_1 and A_2. If the structure is left to stay in this condition for a long time, the potential wells will be filled with electrons arising from thermal generation. The clock rate of CCD operation must certainly be high enough so that the charge of electrons—carriers of information—can be well in excess of the charge of thermally generated electrons.

In this time interval, the CCD records information on the image of the exposed scene. Light-excited photocarriers of one sign (electrons) collect under electrodes A_1 and A_2 in amounts corresponding to the intensity of radiation incident on the CCD surface within the regions of these electrodes. In the transfer (readout) interval t_{tr}, a change in the supply voltages ensures a unidirectional motion of charges from A_1 and A_2 to the output electrode from which video signals are taken off. The amplitude of video signals is proportional to the level of illumination of the corresponding CCD region. Of importance is the fact that an adequate charge transfer necessitates a sufficiently small spacing between electrodes and the surface potential profile free from potential barriers.

From the physics viewpoint, this class of photodetectors is unique in the sense that these devices convert optical radiation (almost always in the visible region) into an electric signal of the specific kind, namely, a signal in the form of charges. The photodetectors considered in the previous sections convert an optical signal into current or voltage.

Optical CCDs offer particular promise as multielement photodetectors. The main distinction between CCDs and other multielement photodetectors is that in the former devices the same elements (structures) pick up, store, and transfer optical information. This feature predetermines the simplicity and low fabrication cost of image sensing CCD arrays.

Chapter 3

Optrons

3.1. Block Diagram of an Optron

3.1.1. Optron properties. The optron is an optoelectronic device consisting of two basic components, namely, a light source and a photodetector optically and structurally coupled together. This device is known as a photocoupled isolator, photocoupler, photon-coupled pair (PCP), photorelay, etc.

Any type of optron operates on the principle of double energy conversion: the radiator converts the energy of an electric signal to optical radiation and the photodetector does the reverse—converts this radiation to electric current or voltage. The optron thus has an electric input and an electric output for electric connection to an external circuit. Inside the device, however, the connection between the input and output is optical: optical signals provide optical coupling (and thus electric isolation) between the radiator and photodetector, hence the name of the device—a photocoupler or photocoupled isolator. Such a device incorporated into an electric circuit performs the function of an output element—a photodetector which simultaneously provides electric isolation (conductive decoupling) between the input and the output signal. But the purpose of the optron does not merely come down to securing electric isolation. The optron can provide optical control of electronic circuits and thus enables the designer to evolve devices with unique parameters and characteristics.

As a circuit component designed to perform definite functions, the optron is primarily defined in terms of what type of photodetector it incorporates. That is why from the application viewpoint, it is the photodetector that represents a decisive element of the optron, the radiator being selected to suit the photodetector. From the design and technological viewpoints, both elements of the optron are equally important. Moreover, in most cases the radiator determines the effectiveness of energy conversion and the life of the optron, so this element is usually given much consideration in designing. Nevertheless, we should stress the point once more—it is the photodetector that determines the circuitry features of the optron.

As mentioned in the introduction to this book, the principal physical advantages of optrons follow from the utilization of photons as carriers of information. These devices feature a very high input-output electric isolation which excludes output-to-input feedback, provide unidirectional flow of information, and have a large bandwidth.

Besides, optrons offer a number of other important advantages:

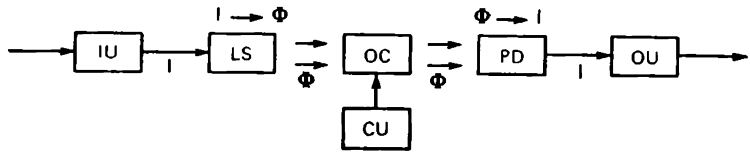


Fig. 3.1. Block diagram of an optron

— contactless (optical) control of electronic devices with the attendant flexibility of the search for design solutions to the problems of control;

— nonsusceptibility of optical channels to the action of electromagnetic fields, which ensures high noise immunity and excludes mutual interference where optrons feature extended optical channels;

— possibility of creating functional microelectronic devices with photodetectors the characteristics of which can vary in any desired manner on exposure to optical radiation;

— control of the optron output signal in a variety of ways by exposing the material of an optical channel to both electrical and nonelectrical influences, which makes it possible to construct diverse types of sensors and devices for information transmission.

The disadvantages inherent in optrons are the following:

— low efficiency due to double energy conversion (electric energy to radiant energy, and vice versa), and appreciable power dissipation;

— strong temperature dependence of parameters;

— high level of intrinsic noise;

— design-fabrication imperfection due, mainly, to the need to employ hybrid technology.

The improvements made in the materials, technology, and circuitry gradually eliminate the above drawbacks. The optrons become very popular due to the unique features of these devices.

3.1.2. Block diagram. The generalized block diagram of an optron and the forms of energy converted in its individual stages are shown in Fig. 3.1. The input unit (IU) serves to convert the input signals to the form suitable for efficient operation of a light source (LS), i.e. radiator, or photoemitter. If an optron is driven from, say, a logic IC, a current of 0.1 to 1 mA drawn from the circuit must be amplified to 10-15 mA. The input unit must also comply with the requirements of economical efficiency and sufficient speed so as not to impair the response of the optron as a whole.

The light source must afford a high efficiency of electrooptic conversion, high speed of response, and a sufficiently narrow beam. Besides, it is commonly desirable that the light source should operate at the lowest input current possible, about 1 mA. For linear systems, it is also of importance to enable the radiator to operate in a wide

dynamic range of input currents; this is the current range in which the quantum efficiency of the light source and thus the current transfer ratio of the optron are held constant.

The optical channel (OC), or optical path, serves to transmit the energy of an optical signal from the light source LS to photodetector (PD) with a minimum of loss and without distortion of the waveform of the optical signal. The channel should secure a minimum of nonforward scattering to exclude the effect of radiation on other light-sensitive elements of the device. It also should have a reliable shield against external radiation that can be responsible for false triggering of the optron.

That the possibility in principle exists for controlling the parameters of the optical channel with the aid of electrooptic or magnetooptic effects is evident from the block diagram, which illustrates a control unit CU for OC. Control over the output signal can be made via both the electric input of the optron and the optical input of the photodetector. It is possible to implement a variety of OC designs with the aim to embody the desired functional capabilities in the optron. Thus an optron with an open OC, i.e. with an air gap between LS and PD, is suitable for reading information from punch tapes or cards moving through this gap. By selecting an optical channel that changes its properties on exposure to external nonelectric influences, the designer can evolve various optoelectronic transducers.

The photodetector must convert an optical signal to an electric signal with a minimum of loss of its information content, for which reason it should have a high responsivity and sufficient speed. A photodetector can sometimes perform the function of a preamplifier too. Operation of the LS-OC-PD chain can obviously be effective only after securing the match between the spectral characteristics of all its elements.

The output unit (OU) shapes the output signal from PD to a standard form convenient for its transfer to the subsequent stages, which are most commonly analog and digital microcircuits or semiconductor switches. As with the input unit, the output unit must be efficient and fast.

Thus for all links of the optron, the efficiency of energy conversion and speed are the main criteria. The elements must also interface adequately. They must match each other in optical characteristics (the LS-OC-PD train), electrical characteristics (IU-S and PD-OU pairs), operating temperature range, service life, mechanical strength, and design-fabrication features. Ensuring the match and compatibility of elements is the central task of optimum design of optrons.

3.1.3. Light source. The basic and most universal photoemitter for optrons can be and indeed is an injection light-emitting diode (LED). The use in future of semiconductor lasers noted for high speed

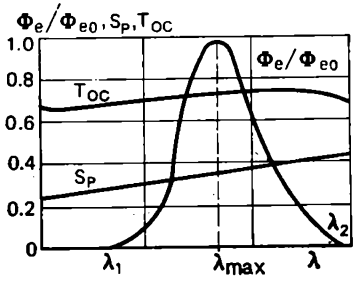


Fig. 3.2. Determining the match between the spectral characteristics of optron elements

and low beam divergence may prove advantageous and economically warranted in the creation of super-high-speed optrons ($t_{sw} \simeq 10^{-9}$ to 10^{-10} s) and high-power optrons.

The optical window of an optron photoemitter differs in design from that of a conventional LED. The LED as a radiator emits from an annular region around a centered bonding pad. This gives a large ratio of apparent-to-actual emitting area. For an optron, the emitting region must be minimum to reduce the edge loss of radiation. The factor that limits a decrease in the emitting region is the permissible current density in the radiator. The bonding pad in an optron photoemitter is offset to allow a minimum of obscuration (shadowing) and to reduce a loss of the flux traveling to PD. A small size of the emitting region also provides for stable optical-coupling conditions, making them practically independent of the accuracy of alignment of the optical window of the photodetector.

3.1.4. Optical medium. The optron performance appreciably depends on the efficiency of energy transfer from LS to PD, i.e. on the properties of the optical channel. To reduce the reflection loss requires the use of an immersion optical system in which the refractive indexes n for the optical medium and the materials of LS and PD are the same in value. This aim is in principle achievable because the values of n in the main types of practical semiconductor materials are close to one another.

Optrons employ the following types of optical medium to provide transparent insulation:

— air, which is a simple means of optical coupling and affords good electric isolation;

— air and lenses to focus the beam, which ensure better radiation transfer as against direct coupling through air;

— an immersion medium, which improves the parameters of OC;

— a light guide, commonly a rigid single fiber cable, which is suitable for use in optoisolators designed to operate at an isolation voltage above 20 to 50 kV.

Thus, in selecting the type of optical channel for an optron destined to perform the required function, it is necessary to fill the following requirements: (1) provide the desired level of input-output isolation (electric isolation between LS and PD); (2) use the OC material that is spectrally fit for LS and PD, i.e. adequately.trans-

parent in the operating wave band; (3) ensure a minimum reflection loss at LS-OC and OC-PD boundaries.

Where the distance between the light source and photodetector is to be small, electric isolation becomes a decisive factor. If use is made of a light guide, isolation requirements are less stringent. But then the transmission spectrum of OC acquires importance since it must be consistent with the spectral characteristic of the LS-OC-PD system (Fig. 3.2).

The measure of spectral consistency of optron elements is the correlation coefficient T_λ for the entire system. This is the ratio between the integrals taken over the products of spectra of the elements forming the optron system:

$$T_\lambda = \frac{\int \frac{\Phi_{e\lambda}}{\Phi_{e0}} S_{P\lambda} T_{OC\lambda}\, d\lambda}{\int \frac{\Phi_{e\lambda}}{\Phi_{e0}}\, d\lambda} \tag{3.1}$$

where $\Phi_{e\lambda}/\Phi_{e0}$ is the normalized spectral characteristic of a photoemitter; Φ_{e0} is the radiant flux at the peak of the emission characteristic, i.e. at $\lambda = \lambda_{max}$; $S_{P\lambda}$ is the relative spectral responsivity of a photodetector; T_{OC} is the relative spectral characteristic of OC; and $T_{OC\lambda}$ is the transmittance of the optical channel.

The integrals are taken between the limits from the short cutoff wavelength λ_1 to the long cutoff wavelength λ_2 of the photoemitter spectrum.

In the general case, the material of OC differs from those of LS and PD in refractive index n. It is known that, as optical radiation propagates from one medium with refractive index n_1 into another medium with n_2, a fraction of radiation is lost as it reflects backward from the interface. This radiation loss is known as the Fresnel loss. The fraction reflected backward is equal to

$$R_{Fr} = \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 \tag{3.2a}$$

Since the difference between n_2 and n_1 enters into the formula as the square, the Fresnel loss is independent of whether the refractive index of one medium is higher or lower than that of the other into which the radiation propagates. The fraction transmitted across the interface is

$$T = 1 - R_{Fr} = 1 - \left(\frac{n_2 - n_1}{n_2 + n_1}\right)^2 = \frac{4n_2 n_1}{n_2^2 + 2n_1 n_2 + n_1^2} \tag{3.2b}$$

The light output efficiency η accounting for the Fresnel loss is the quotient of the numerator and denominator of (3.2b) by the product $n_2 n_1$:

$$\eta = \frac{4}{2 + n_2/n_1 + n_1/n_2} \tag{3.2c}$$

Thus, for a GaAsP source with $n_1 = 3.4$ that emits directly into the air ($n_2 = 1$), the value of η is equal to

$$\eta = \frac{4}{2+1/3.4+3.4/1} = 0.702$$

This means that 70.2% of radiation reaches the photoemitter crystal surface and passes through the interface between the crystal and the air which acts as an optical channel.

The Fresnel loss can be reduced to increase η if the surfaces of LS and PD are given an antireflection coat with a suitable refractive index. Such a coat also affords electric isolation since it is a dielectric material.

For an optron having a GaAsP emitter ($n_1 = 3.6$) and silicon PD ($n_2 = 3.5$) and using the air as a coupling medium ($n = 1$) in one case and a plastic antireflection coating ($n = 1.5$) in the other case, we get

$$\eta = \eta_1\eta_2 = \frac{4}{2+3.6/n+n/3.6} \frac{4}{2+n/3.5+3.5/n}$$

$$= \begin{cases} 0.470 \text{ for } n=1 \\ 0.698 \text{ for } n=1.5 \end{cases}$$

The antireflection coat thus reduces the Fresnel loss by a factor of about 1.5.

3.1.5. Photodetector. The most popular types of PD used in modern optrons are *pin* photodiodes, phototransistors, and photothyristors. The spectral characteristics of these PDs cover the entire visible region and a portion of the near IR region.

Silicon photodiodes are good detectors but need amplifiers to produce the required signal levels. To add gain by use of elements external to the optron package is an inconvenient approach since it leads to a larger size of the system. It is more advantageous to include an amplifier within the optron package. This can be done by two methods. The first is an hybrid method which permits separate optimization of the photodiode and its amplifier and, hence, provides good optical and electrical parameters of the device. The second is an integrated method which reduces the assembly cost but brings a compromise on parameters.

There are two ways to integrate the photodetector and amplifier: the use of a phototransistor in which the collector-base junction acts as a radiation detector; the use of a photodiode in combination with a transistor formed on the same chip to amplify the photocurrent. It is easier to fabricate a phototransistor than a photodiode/transistor, but the former is slower in speed. Photodarlingtons and photothyristors also enjoy wide use in optrons.

3.1.6. Optron classification. There are a number of designs of LS-PD pairs in which electrooptical compatibility of elements is

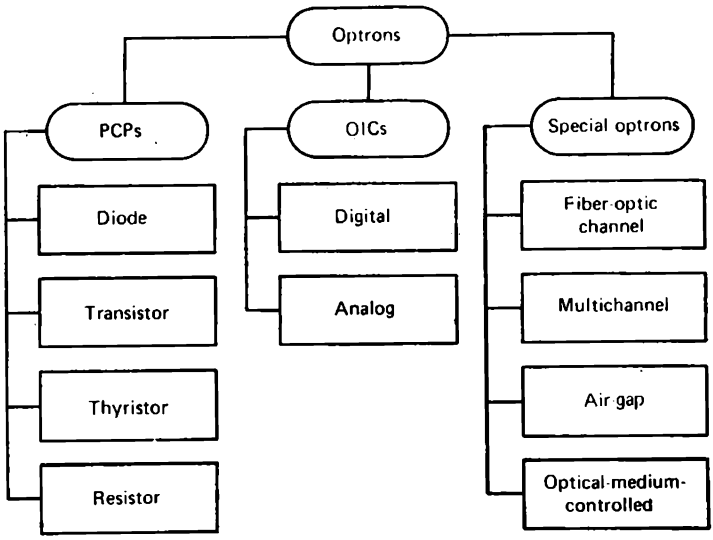


Fig. 3.3. Classification of optrons

achievable to a higher or lesser degree. At present, the following combinations can be considered optimal.

1. Silicon *pin* photodiodes and GaAs(Zn), GaAlAs, GaAsP, and GaAs(Si) emitters. The photodiode matches well the first three types both in spectral response and in speed, and the last type in spectral response.

2. Silicon phototransistors or photothyristors and GaAlAs and GaAs(Si) emitters. The spectral match of the PD with the first type of radiator is good, but somewhat poorer with the second. Here the PD in essence determines the speed of the optron.

3. Photoresistors of the CdS and CdSe types and GaP and GaAsP emitters, which are well compatible in spectral characteristics, but provide a low speed.

These are the main source-detector pairs, which form the basis for modern optrons designed to serve various purposes. The correlation coefficient T_λ for these pairs ranges from 0.6 to 0.9.

By the complexity of block diagrams and the types of optical medium employed, optrons can be broken down into the following groups.

1. Photocoupled (photon-coupled) pairs (PCPs), each consisting of a photoemitter and photodetector optically coupled together and thus electrically isolated from each other.

2. Optronic (optoelectronic) integrated circuits (OICs), each containing one or a few PCPs.

3. Special optrons each consisting of an emitter (emitters) and a

photodetector (photodetectors) with a specific design of the optical channel.

This classification of modern optrons is shown in Fig. 3.3. The type of photodetector determines the classification of photocoupled pairs, so there are resistor, diode, transistor, and thyristor optrons.

A particular optron may thus include all or some of the elements illustrated in the general block diagram of Fig. 3.1. The photon-coupled pair is an LS-OC-PD system. The OIC commonly includes an output unit (OU) and sometimes an input unit (IU) in addition to the above combination of elements. Special optoisolators differ from PCPs and OICs in design, functions performed, and the type of coupling medium used.

3.2. PCP Parameters

The system of parameters of photon-coupled pairs can be broken into four groups: input (I) parameters; output (O) parameters (PD parameters); signal transfer parameters; and isolation (I/O) parameters.

In the text below, we shall use the following abbreviations for the basic types of PCP: D (diode), T (transistor), R (resistor), Y (thyristor), T^2 (photodarlington), and DT (photodiode/transistor) optrons (Fig. 3.4).

3.2.1. Input and output parameters. The input circuit of a PCP is defined by the following basic parameters (commonly for LEDs).

1. Nominal input current $I_{I\,\mathrm{nom}}$. This is the rated value recommended for optimum operation and used in the measurement of other basic parameters of PCPs.

2. Input voltage V_I, which is the voltage drop across a forward biased LED at the specified value of forward current, commonly $I_{I\,\mathrm{nom}}$.

3. Input capacitance C_I between the input terminals of a PCP operating in specified conditions.

The limiting input parameters are the following.

1. Maximum input current $I_{I\,\mathrm{max}}$. This is a maximum permissible value of direct forward current which is still allowed to be passed through the optron LED.

2. Reverse input voltage V_{IR}. This is a maximum permissible value of reverse voltage of any form (dc, pulse, sinusoidal, etc.) that can be applied to the PCP.

The output parameters of photocoupled pairs, i.e. the parameters of photodetectors, are listed in Table 3.1.

3.2.2. Transfer parameters. Specifically optronic parameters include:

— direct coupling parameters, which characterize the effect of the light source on the photodetector and the attendant features

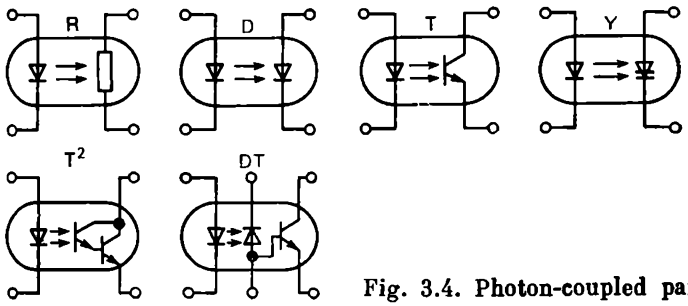


Fig. 3.4. Photon-coupled pairs

peculiar to the transfer of a signal from the input to the output circuit of the PCP as an optically coupled isolator;

— isolation parameters (feedback parameters), whose values are indicative of how much the optocoupler approaches an ideal isolator.

Table 3.1. PCP Output Parameters

Parameter	Definition	PCP type
Maximum permissible reverse output voltage, V_{OR}	Maximum value of reverse voltage of any form allowed to be applied to PCP output circuit	T, T², DT, D, Y, R
Maximum permissible output current, $I_{O\,max}$	Maximum value of current allowed to flow through PD with PCP switched on	T, T², DT, Y, R
Output leakage current, I_{lk}	Output current at $I_I=0$ and the specified value and sign of V_O	T, T², DT, D, Y
Output light resistance, R_l	Value of photoresistor at specified input current	R
Output dark resistance, R_d	Value of photoresistor at $I_I=0$	R
Output saturation (residual) voltage, V_{sat}	Voltage across photothyristor or phototransistor in saturation	T, T², Y
Output capacitance, C_O	PD capacitance	Any type

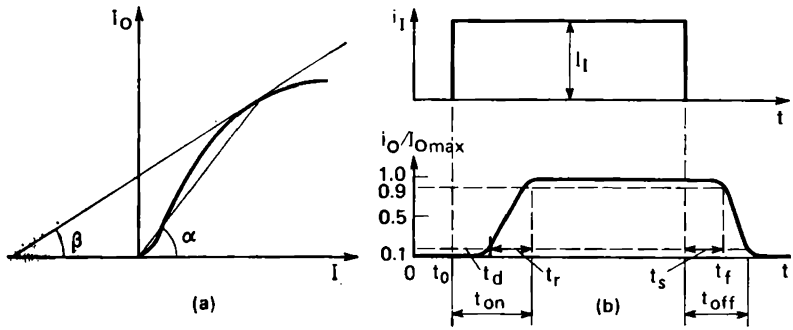


Fig. 3.5. Determining the static and the differential CTR (*a*) and waveform describing transients in a PCP (*b*)

The basic parameter of all PCPs, except for the thyristor type, is the current transfer ratio (CTR), labelled K in formulas, figures, and tables for brevity, which characterizes the signal transfer from the input to the output circuit of a PCP. This is the gain defined as the ratio of the output current to the input current:

$$K = I_O/I_I \tag{3.3a}$$

Where the leakage current I_{lk} at the PD output is comparable in magnitude with the output current, for example, in operation at high temperature, a correction for I_{lk} should be introduced into (3.3a) to calculate the CTR

$$K = (I_O - I_{lk})/I_I \tag{3.3b}$$

The CTR for optoelectronic devices is usually given in percent and listed in specifications for D, T, T^2, DT, and R photon-coupled pairs.

The I_O-I_I relation, which is known as the transfer characteristic, is generally nonlinear for PCPs. Formulas (3.3a) and (3.3b) define the static CTR. The differential CTR is given by

$$K_{dif} = \lim_{\Delta I_I \to 0} \Delta I_O/\Delta I_I = dI_O/dI_I \tag{3.3c}$$

The static CTR is proportional to the slope α of a straight line drawn from the origin to the point of intersection on the I_O-I_I transfer curve (Fig. 3.5*a*), i.e.

$$K \sim \tan \alpha$$

The differential CTR is proportional to the slope β of the straight line tangent at the given point to the transfer curve, i.e.

$$K_{dif} = \tan\beta$$

The nonlinearity of the transfer characteristic tells us that the current transfer ratio varies with the input current. This is because the

photoemission characteristic deviates from the linear pattern and the amplifier gain varies with the output current.

For transistor PCPs operating in the switch mode, the parameter K determines the load capability of the device: if the load connected to the output port of the PCP is such that the output current I_O does not exceed the value KI_O, the output transistor switch will be on in saturation. The voltage drop across the output transistor will be equal to the voltage between the collector and emitter of the saturated transistor, V_{sat}. When $I_O > KI_I$, the condition of saturation no longer holds, thus disturbing the transistor switch operation.

The parameter K for resistor PCPs is of a formal character: in place of I_O, expression (3.3a) then includes the photoresistor current corresponding to the end of the linear segment of the photoresistor VI curve. The transfer characteristic of an R PCP is practically a function of the input current which yields a specified value of the output parameter such as a light resistance R_l or light-to-dark resistance ratio R_l/R_d.

A thyristor PCP principally operates only in the switch mode, and the parameter K for this device has no meaning at all. The transfer characteristic here is defined by the input current at which the photothyristor turns on. The certificates on Y PCPs list the value of switching current, which is a minimum value of input current enabling the thyristor to be reliably driven into conduction. Besides, it specifies a maximum permissible input noise current $I_{N\,max}$ at which Y still stays off. The parameter $I_{N\,max}$ characterizes the noise immunity of thyristor PCPs.

The speed of response of PCPs depends on the following parameters of the transient characteristic which describes the output current $i_O(t)$ as a function of the current input I_I squared (Fig. 3.5*b*):

— turn-on delay time t_d, which is the time required for the output current to reach 10% of $I_{O\,max}$ from the moment t_0 at which the input current has commenced;

— rise time t_r, which is the time it takes for the output current to rise from 10% to 90% of $I_{O\,max}$.

The turn-on time t_{on} for a PCP is the sum of t_d and t_r, $t_{on} = t_d + t_r$. The waveform of Fig. 3.5*b* that illustrates the transient in a PCP also shows the storage time t_s and fall time t_f at switch-off and also the turn-off time t_{off}. The sum of the turn-on and the turn-off time is the switching response time $t_{sw} = t_{on} + t_{off}$, which gives the total estimate of the speed of a photocoupled pair.

The speed of some classes of PCPs is sometimes specified by the cut-off frequency f_c or maximum rate of data transfer, both being related to the value of t_{sw}.

3.2.3. Isolation parameters. These include a peak permissible in/out isolation voltage V_p, maximum permissible static in/out

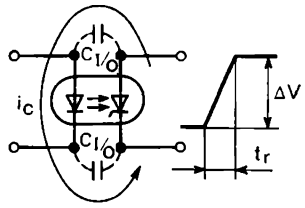


Fig. 3.6. The effect of transfer capacitance $C_{I/O}$ in a PCP

isolation voltage $V_{I/O}$, in/out isolation resistance $R_{I/O}$, transfer (in/out) capacitance $C_{I/O}$, and maximum permissible rate of rise of the output voltage, $(dv/dt)_{max}$.

The most important parameter is V_p; it is precisely this parameter that determines the isolation capabilities of a photocoupled pair. The voltage V_p is the peak value of voltage of the given duration and repetition rate that can be held between the input and the output port without degrading the dielectric strength of the PCP. The static in/out isolation voltage $V_{I/O}$ determines the breakdown strength of a PCP at dc voltage (in the static mode, which calls for presetting the decoupling resistance at $V_{I/O}$). The parameters $C_{I/O}$ and $(dv/dt)_{max}$ define the PCP stability to output voltage surges. Where these surges are present and the rate of voltage rise is high, a capacitive current i_c may pass through the PCP photoemitter (Fig. 3.6):

$$i_c = \frac{1}{2} C_{I/O} \frac{\Delta V}{t_r} \tag{3.4}$$

where $C_{I/O}$ is the transfer input/output capacitance.

An excessively large current i_c may cause false triggering of the PCP even if the voltage swing, V_{max}, is much lower than V_p.

Photocoupled pairs virtually eliminate dc feedback because the isolation resistance $R_{I/O}$ reaches 10^{12} ohms. But the ac ground loop current can be appreciable since the value of transfer capacitance is about 1 pF for most PCPs. The only recourse made at present to minimize $C_{I/O}$ is to increase the separation between the input and output circuits of PCPs. But such PCPs are usually costly because they require lenses or fiber optics to retain the efficiency at an adequate level.

3.3. Basic Types of Optrons

3.3.1. Diode photocoupled pairs (D PCPs). A typical representative of this class of PCPs is an optoisolator comprising a GaAs emitter and silicon *pin* photodiode with a silicon junction coat as an optical medium (optical channel) sandwiched between these two elements.

In operating conditions, this type of PCP cannot practically afford a switching time t_{sw} of 20 to 40 ns, as listed in the certificates on these devices. Indeed, the PCP load operating into a subsequent stage must be 1-10 kilohms at the small output currents provided by this PCP. Since the output self-capacitance of the PCP is about 40 pF, the recharge time constant is as high as 0.1-1 μs.

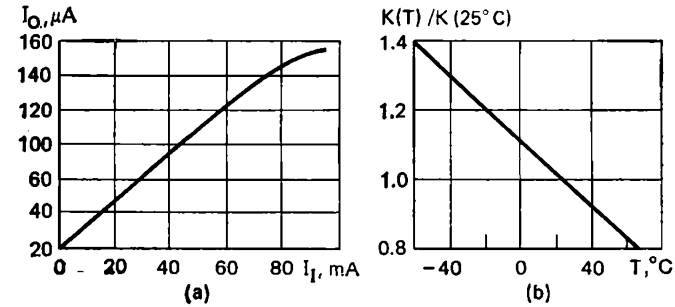


Fig. 3.7. Transfer characteristic of a diode PCP (*a*) and CTR versus temperature relation (*b*)

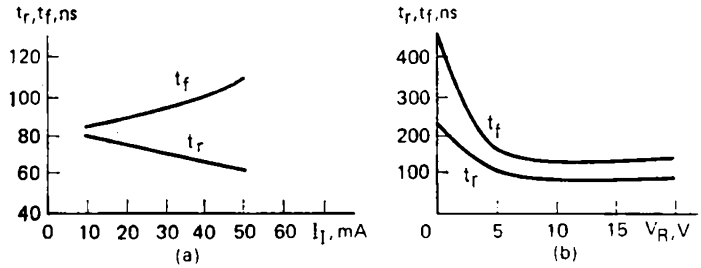


Fig. 3.8. Time-varying parameters of PCPs as functions of input current (*a*) and reverse voltage (*b*)

Thus, the speed of a D PCP is most commonly dependent on its output capacitance. Manuals cite PCP speeds rated at a load of 50 ohms, which in fact characterizes only the upper speed limit.

Diode PCPs feature a good linearity of the transfer characteristic (Fig. 3.7*a*); the normalized change of the current transfer ratio K does not exceed $\pm$ 0.05 in the range of input currents from 20 to 80 mA.

A disadvantage of these optoisolators is a strong temperature dependence of K (Fig. 3.7*b*) due, mainly, to variations in the photoemitter parameters with temperature T; in the temperature range from −60 to + 85°C, the K-T relation is practically linear and the temperature coefficient of K, % °C^{-1}, is equal to

$$\frac{dK/dT}{K} \approx -0.5$$

The plots of dynamic parameters of photocoupled pairs versus input current and reverse voltage appear in Fig. 3.8. A change in the input current has a different effect on the values of t_r and t_f (Fig. 3.8*a*); t_{sw} thus practically remains invariable. It is characteristic that even at reverse voltages V_R of 3 to 5 V, the transients in the *pin* diode proceed so fast that the PCP speed only depends on the

photoemitter speed and is independent of the diode voltage (Fig. 3.8*b*).

The class of diode photocoupled pairs includes the optoisolators (OIs) proper, in which transparent glass or any other properly encapsulated light conductor 40 to 100 mm long provides both optical coupling and electric insulation between the light source and detector. Such a structure reduces the transfer capacitance to minute values of about 0.01 pF and increases the dielectric strength to 20-50 kV.

Thus, diode PCPs offer low values of the current transfer ratio (K = 0.5 to 20%), relatively high speed (t_{sw} = 10 to 30 ns), and, as a rule, high dielectric strength (2-7 to 20-50 kV for OIs using transparent junction coating).

Diode photocoupled pairs reveal the level of development of optoelectronic engineering more vividly than any other class of devices. By the values of K we can judge of the attained efficiency of energy conversion in a PCP; the values of dynamic parameters allow us to assess the limiting speed of the entire class of optrons.

3.3.2. Transistor photocoupled pairs (T PCPs). These devices form the most representative class of optrons. The popularity of transistor PCPs comes from their ability to provide output current sufficient to drive subsequent stages.

The photodetector here is commonly a silicon planar *npn* transistor with the base left floating or also connected to the external circuit. The light source is usually a GaAs LED.

The characteristics of a transistor PCP substantially differ from the analogous characteristics of a diode PCP. The transfer characteristic deviates from the linear form. This deviation rises with gain of the transistor and input current. Fig. 3.9 illustrates the plots of the CTR versus temperature. At high input current (curve *2*), the K-T relation is the same as for a diode PCP, but differs appreciably at small currents (curve *1*). The shape of the curves is illustrative of the manner in which the emissive power of the radiator and the gain of the phototransistor vary with temperature and current.

A photodarlington used in the T PCP strongly increases the CTR, but at an appreciable sacrifice in speed of response and temperature stability. With a change in temperature from 25 to 100°C, the dark current through the photodarlington increases 40- to 50-fold as against a 20- to 30-fold increase in an ordinary phototransistor. This circumstance places a reduced upper limit on the operating temperature of PCPs.

The main parameters characterizing the efficiency of the above described PCPs as linear converters are the CTR, switching time t_{sw}, and isolation voltage $V_{I/O}$ which defines the ability of a PCP to provide electric insulation between the input and output circuits. These figures of merit are interdependent: a change (improvement)

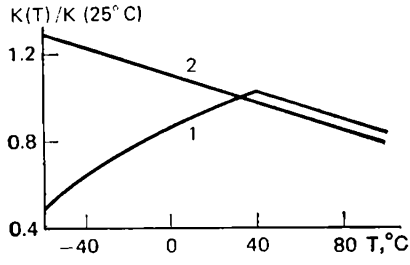


Fig. 3.9. CTR of a transistor PCP versus temperature

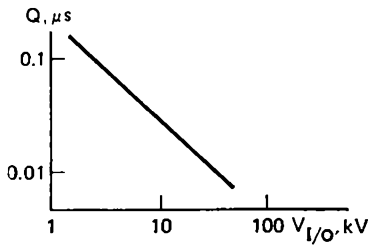


Fig. 3.10. Relation between Q factor and isolation voltage

of one parameter commonly leads to a change (impairment) of another. For example, an increase in $V_{I/O}$, which necessitates a large separation between LS and PD, degrades optical coupling and reduces the K value.

One more figure of merit of PCPs is the quality factor Q defined as

$$Q = K/t_{sw}$$

For D, T, and T^2 photocoupled pairs, the Q factor is practically constant since attempts made to increase the CTR lead to a higher t_{sw}, and vice versa.

Figure 3.10 presents an experimental plot of dielectric strength versus Q for diode, transistor and photodarlington (T^2) PCPs. As seen from the graph, the Q for all types of PCP at a specified $V_{I/O}$ is approximately constant. At the present level of development of incoherent optoelectronics, the Q ranges between 0.1 and 1 μs for typical values of $V_{I/O}$ from 1 to 5 kV. The need to increase the Q value of PCPs is one of the main tasks of optoelectronics today.

3.3.3. Thyristor photocoupled pairs (Y PCPs). These devices used as power switches have good electric isolation between the control and the anode circuit. Thyristor PCPs are finding ever increasing use today; this type of optoisolator will probably be the basic element in high-power converters. The input circuit of a thyristor PCP that can deliver large powers of up to 100 kW to a load is practically compatible with integrated circuits.

3.3.4. Resistor photocoupled pairs (R PCPs). As mentioned earlier, the resistor PCP does not take advantage of a high speed of the photoemitter driving the photoresistor. This device cannot provide an optimum match in either spectral or electrical characteristics. The need to improve the parameters of resistor PCPs and primarily to increase the speed has led to the development of PCPs with *pn* junction photodetectors.

3.3.5. Optoelectronic integrated circuits (OICs). At present OICs are mainly made by semiconductor hybrid technology, so they be-

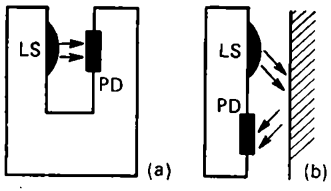


Fig. 3.11. Optrons using open optical paths
(*a*) chopped-beam responsive;
(*b*) reflected-beam responsive

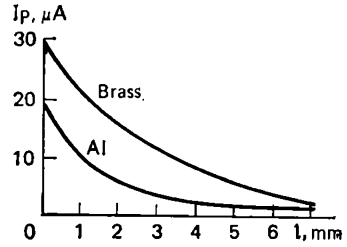


Fig. 3.12. Output current of a reflected-beam responsive optron as a function of distance to reflecting surface

long to hybrid ICs. Structurally, the OIC consists of an unpackaged IC and unpackaged PCP, mostly a diode optoisolator. OICs fall into two groups, digital OICs and analog OICs depending on the type of IC used. At present, OICs have limited applications because of a relatively low speed of available optoisolators. The only type of optoisolator that is adequately compatible with ICs in speed is a diode PCP. It also offers the best linearity of its transfer characteristic and is the most promising device for use in analog applications. Along with diode PCPs, differential optoisolators hold promise in analog applications.

The discussion of microcircuits of integrated optics (elements of micro-optoelectronics) that use an optical signal as the only carrier of information is outside the scope of this book.

3.3.6. Special optrons. These optoisolators differ from traditional photon-coupled pairs and optically coupled ICs by the type of optical medium they use to couple the LS to PD.

Optrons using an open optical medium. In these devices the air gap between the LS and PD serves as an optical path for transmitting an optical signal to the photodetector. Any mechanical obstacle placed in the air gap will change the intensity of the beam and, hence, the amplitude of the output signal from the PD. These optrons perform the functions of optical sensors used to discriminate objects, to assess the status of a surface, to measure the rate of motion, etc. Fig. 3.11 illustrates two design versions of these sensors; the first (Fig. 3.11*a*) is a chopped-beam responsive isolator and the second (Fig. 3.11*b*) is a reflected-beam responsive isolator. The CTR here is obviously small due to the air gap. These devices usually employ a phototransistor in the output circuit. In the optical sensor of the second type, the CTR depends on the distance at which an object lies in front of the PD and on the properties of the object surface. Typical plots of output current versus distance from the reflected-beam responsive optron to the reflecting surface are shown in Fig.

3.12. It is seen that the sensor discriminates an object at a distance of up to several millimeters. The operating frequencies of these devices cover the range of up to 1 MHz.

Air-gap optrons can also act as detectors of gas or moisture pollution of environments. A detector contains two light sources, one of which emits at the wavelength corresponding to a maximum of absorption of light by the gas contaminant and the other source serves as a reference. Most popular are optronic sensors of humidity of various media ($\lambda_{max} \approx 1.6$ μm).

Fiber-optic channel optrons use a fiber-optic conductor (light guide) from tens of cm to tens of meters long for optical coupling between the LS and PD. Such an optical path provides loss-free transmission of optical signals as well as conductive decoupling. The optron here acts as a short communication line which, unlike a wire communication line, offers the following advantages inherent only in optrons: a decreased size and mass due to a lower mass density of the materials used and freedom from heavy sheathing; good noise immunity of the channel; and leak-free transmission of information since radiation cannot propagate into the environment. Optical pulses transmitted through the light guide suffer an appreciable loss. That is why the CTR here is very small, less than 0.01%. Besides, these devices require far more complex matching means than optron ICs.

3.3.7. Parameters and characteristics of industrial photon-coupled pairs. In conclusion, consider the parameters and characteristics of PCPs commercially available in the USSR.

The analysis of parameters of diode PCPs shows that these devices can be broken down into several groups, as illustrated in Table 3.2. General-purpose D PCPs have K values ranging from 1 to 1.5%, and t_{sw} values from 100 to 300 ns. This group of PCPs features a wide range of output voltages suitable for a variety of devices.

PCPs designed to operate in the photovoltaic mode are noted for increased K values, above 3 or 4%. In contrast to the CTR of other PCPs, the CTR for this group of devices is measured in the most stringent test conditions without applying a reverse voltage on the photodetector.

Unpackaged PCPs intended for use in hybrid microcircuits and microassemblies (VLSICs) generally have low values of the reverse voltage across the photodiode. Note that the fastest diode photocoupled pairs, whose t_r is lower than 30 ns, enable a very high rate of data transfer, up to 15-20 Mbit per second.

The main parameters of transistor PCPs commercially produced in the USSR are listed in Table 3.3. The comparative analysis of these devices permits us to conclude the following:

(1) diode-transistor, transistor, and photodarlington (T^2) PCPs differ widely in the values of basic parameters;

Table 3.2. Diode PCPs

Type	K	t_r, ns	$I_{I\,nom}$, mA	V_I, V	$V_{I/O}$, kV	$R_{I/O}$, Ω	$C_{I/O}$, pF	Note on PCP type
A	0.010	100		1.5				
B	0.015	500		1.5				
C	0.012	1000	10	1.5	0.2-2.0	10^9	2	General-purpose
D	0.007	500		1.5				
E	0.010	250		1.8				
A	0.006-0.013	100						
B	0.009-0.020	100						
C	0.015-0.035	100						
D	0.006-0.013	800	10	1.5	0.1-0.2	10^{10}	1.8	Unpackaged, for digital applications
E	0.009-0.020	800						
F	0.015-0.035	800						
A	0.010	30	10	1.7	0.4-1.0	10^{10}	2	High-speed, unpackaged
B	0.006							
A	0.025	3000	20	1.7	0.1-0.2	10^{10}	2.5	Unpackaged, operating in PV mode
A	0.015	100	10	1.7	0.2-0.4	10^{10}	1	General-purpose, unpackaged
B	0.025	150				10^9	2	

(2) DT PCPs are the fastest and T^2 PCPs are the slowest;

(3) T^2 PCPs have the highest CTR, 100 and above, and the lowest output currents and voltages.

The CTR versus input current relation for a transistor operating in the active region of its characteristic does not exhibit a distinct portion where the CTR is constant (unlike that for D PCPs) because the β value of the transistor depends on I_b.

The plots of dynamic (time-varying) parameters versus input current (Fig. 3.13) take on the same character as for diode PCPs; in evidence is a sharp rise in t_d and t_s at small values of I_I. This behavior is common to all photocoupled pairs using LEDs. As the temperature grows, the speed of transistor PCPs drops off; this trend is particularly conspicuous at switch-off for the parameters describing the trailing edge of the pulse (Fig. 3.14).

Thyristor photocoupled pairs (Y PCPs) possess the geatest promise for use as contactless switches of high-current high-voltage circuits: regarding the powers they are able to switch into a load in combina-

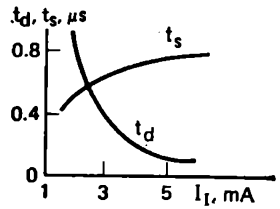


Fig. 3.13. Time-varying parameters of diode-transistor PCPs versus input current

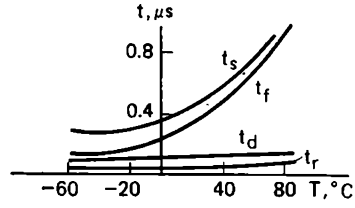


Fig. 3.14. Time-varying parameters of diode-transistor PCPs versus temperature

tion with the speed of response, these PCPs are clearly more preferable than T^2 PCPs. Medium-power Y PCPs are meant for jobs in various electronic circuits such as control devices, power amplifiers, and pulse shapers. Thyristor PCPs are divided into several groups depending on the rated values of currents and voltages they are able to switch (Table 3.4).

Table 3.3. Transistor PCPs

Type	K	t_{sw}, μs	$I_{I\ nom}$, mA	V_I, V	I_O, mA	V_{sat}, V	$V_{I/O}$, kV	$R_{I/O}$, Ω
Diode-transistor								
A	0.1	2.0						
B	0.2	3.0	10	1.5	5.0	—	0.2-3	10^{10}
C	0.4	4.0						
Transistor								
A	0.5	8.0-30		1.5	5.0	0.4		5×10^{8}
B	0.5	8.0	20	2.0	10	0.3	0.2-3	10^{9}
C	1.0	8.0		2.0	20	0.5		10^{9}
Photodarlington								
A	8	150	25	2.0	200	1.5	0.2-3	10^{9}
B	4				100			

High-power thyristor PCPs are designed for work in dc and 500-Hz ac power circuits. The devices are mounted in flat plastic packages or metal-glass cans. These PCPs use solely LEDs with a relatively low values of permissible forward current, for which reason the input circuits of the PCPs mainly operate in the pulse mode. Power Y PCPs fall into a large number of groups depending on the maximum values of output current and output voltage, switching time, maximum permissible rates of rise of forward voltage and forward cur-

Table 3.4. Thyristor PCPs

Parameter	PCP type	
	low-power	power
Output current $I_{O\,max}$, A	0.1	6.3-320
Output forward voltage $V_{O\,max}$, V	50-200	100-1300
Saturation voltage V_{sat}, V	1.8	1.4-1.85
Input voltage V_I, V	1.5-2.0	2-3
Input current $I_{I\,max}$, mA	55	100
Nominal input current $I_{I\,nom}$, mA	20-50	50-200
Turn-on time t_{on}, μs	10	10-30
Turn-off time t_{off}, μs	35	50-250
Isolation voltage $V_{I/O}$, V	500	1000-3000
Isolation resistance $R_{I/O}$, Ω	5×10^8	—
Transfer capacitance $C_{I/O}$, pF	3	—
Input noise current $I_{N\,max}$, mA	0.5	—
Output voltage rise rate $(dv/dt)_{max}$, V/s	5×10^6	(20-100) 10^6

rent (20 to 200 A/μs). In power PCPs, the LED emits light to the photodetector through an air gap, so the values of $R_{I/O}$ and $C_{I/O}$ are exclusively dependent only on the package parameters.

The basic parameters of resistor photocoupled pairs are given in Table 3.5.

There are fairly many types of resistor PCPs, of which the most popular are the first three types designed to switch high-voltage circuits, in particular, to control electroluminescent displays. All

Table 3.5. Resistor PCPs

PCP type	Emitter type	Input parameter		Output parameter			Output resistance		t_{sw}, ms
		V_I, V	I_I, mA	V_O, V	I_O, mA	P_O, W	R_l, Ω	R_d, Ω	
1	Filament lamp	5.8	16	250	3.5	0.05	$2\cdot10^3$	$3\cdot10^8$	200
2	Filament lamp	5.8	16	250	7	0.08	$4\cdot10^2$	$3\cdot10^7$	200
3	LED	3.8	15	250	3.5	0.04	$1.2\cdot10^3$	$3\cdot10^8$	150
4	Filament lamp	5.8	16	20	0.2	0.025	10^4	10^9	100
5	Filament lamp	5.8	16	20	0.2	0.025	10^6	10^{11}	100
6	Filament lamp	5.8	16	250	2	0.025	$4\cdot10^2$	$1.5\cdot10^7$	200
7	Filament lamp	5.8	16	250	2	0.025	$3\cdot10^3$	$1.5\cdot10^8$	200
8	Filament lamp	5.8	16	10	1	0.025	$1.5\cdot10^2$-10^3	10^7	200
9	Filament lamp	5.8	16	10	1	0.015	$1.5\cdot10^2$-10^3	10^7	200
10	LED	3.8	10	35	0.2	0.01	$2\cdot10^3$	10^6	120
11	LED	3.8	10	35	0.2	0.01	$2\cdot10^3$	10^6	120
12	LED	2.5	10	5	—	0.005	10^3	10^7	0.5

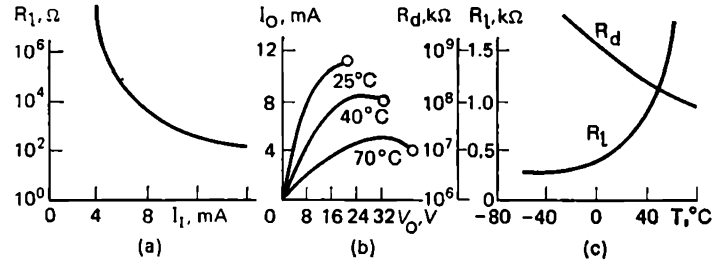


Fig. 3.15. Characteristics of resistor PCPs
(*a*) plot of light resistance versus input current; (*b*) *VI* curves; (*c*) plots of dark resistance and light resistance versus temperature

types are suitable for use in linear circuits, but types 9, 10, and 11 are most convenient since they allow for compensation of the temperature and time-dependent drifts of parameters.

A search for the ways of increasing the speed of resistor PCPs is one of the key tasks in this domain. The use of CdS-ZnS films enabled a reduction of the relaxation time to units of milliseconds. Photodetectors based on single crystals of Si doped with zinc also gave hopeful results, and CdSe-HgSe photosensitive layers made possible a commercial yield of relatively fast R PCPs.

The analysis of typical relations for the parameters of resistor PCPs allows us to draw the following conclusions.

1. Wide variations of the resistance in the range from 4 to 8 orders of magnitude occur in a rather narrow interval of input currents (Fig. 3.15). A relatively high value of light resistance R_l causes a noticeable self-heating, for which reason the *VI* curves deviate from the linear run at output currents of several milliamperes (Fig. 3.15*b*). A further increase in the output voltage gives rise to negative resistance which places an upper limit on the power dissipation in the photoresistor.

2. The parameters of resistor PCPs heavily vary with temperature (Fig. 3.15*c*). A typical feature is that R_l rises with temperature at a rate of about 1.5% °C, while the dark resistance R_d sharply drops so that the range of changes in R_d/R_l decreases and may fall at T_{am} = 60 to 70°C to an impermissibly low value of 10^2 to 10^3. On cooling, the time lag of the photocoupled pair grows very appreciably, and the switching time may reach a few seconds. Large values of t_{sw} also limit the operating frequency range of photocoupled pairs: only the 'fastest' PCPs can operate at up to 1000 Hz, while for other PCPs the frequency range does not exceed tens of hertz.

3.4. Application of Optrons

3.4.1. Classification of optrons by the fields of application. The analysis from this viewpoint allows us to single out three basic classes of optrons.

1. Digital (pulse) optrons intended for high-speed transfer of digital information between electrically isolated circuits. The figure of merit for this class of optrons is the quality factor Q defined by (1.21). This parameter indeed determines the ability of an optron to transfer a digital signal: an increase in the CTR is conducive to a lower loss in the pulse amplitude, shorter t_{sw}, and smaller distortion of the leading and trailing edges of the pulse. Regarding the isolation parameters of these optrons, $C_{I/O}$ is the decisive parameter because parasitic feedback is most appreciable in the dynamic mode of operation of high-speed devices.

2. Analog (linear) optrons designed for undistorted transmission of analog signals between conductively decoupled circuits. The requirements for analog optrons vary according to particular applications and are rather specific. That is why the quality factor used for digital optrons is unacceptable here. Of the isolation parameters, $R_{I/O}$ is most important.

3. Power switches intended for contactless control of high-voltage power circuits. The generalized parameter of this type of switch is the switching Q factor:

$$Q_{sw} = P_{sw}/P_I \tag{3.6}$$

where P_{sw} is the maximum power of the switched circuit; and P_I is the input power required for circuit control.

In the case of high-speed switching, the Q factor is:

$$Q_{sw} = P_{sw}/P_I t_{sw}$$

For modern optrons, $Q_{sw} = 10^2$ to 10^6. The parameter $V_{I/O}$ in the main determines the isolation properties of an optron switch.

The class of digital optrons essentially includes diode and transistor optrons; as for analog optrons, these can be diode, resistor, and, sometimes, transistor optrons, but multichannel (differential) types look most promising for the purpose. Typical examples of optronic power switches are thyristor and transistor optrons (mainly photodarlington types).

3.4.2. Application of optrons in digital and pulse devices. Comparison between the dynamic parameters of various optrons and analogous parameters of discrete diodes, transistors, and microcircuits reveals that the optrons are at present inferior in speed to electronic counterparts (see Table 3.1). A decrease in speed of the circuit incorporating an optron proves comparatively small if the operating conditions of optron elements are made consistent. The use of LEDs

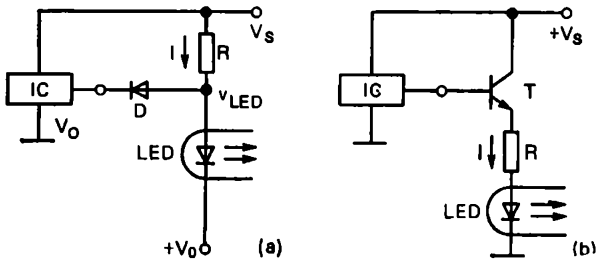


Fig. 3.16. Matching the optron input to the IC output
(*a*) circuit with resistor R to specify input current; (*b*) circuit with transistor amplifier to provide match in input current

and diode or transistor photodetectors makes the speed of optrons mainly dependent on the recharge of depletion capacitances of the photoemitter and photodetector. Thus, the typical values of capacitances for photoemitters of high-speed optrons reach 20-300 pF; for photodiodes and phototransistors these values lie between 5 and 15 pF. Therefore, for the switching time of an optron to be reduced, it is necessary (1) to speed up the recharge of the photoemitter input capacitance and (2) to diminish the recharge time of the photodetector output capacitance by isolating or compensating for the capacitive load and decreasing the output voltage swing.

Control of optrons by appreciably high currents enables fast switching of a light source with the simultaneous generation of a powerful and stable flux. For this we should, first, clamp the forward current of the LED to ensure a stable flux and, second, keep the LED current below the upper permissible limit. It is also of importance that the supply circuits of LEDs should be versatile to make the LED performance invariable after replacing one of the elements of the circuit.

Consider, for example, how an integrated circuit switches a LED (Fig. 3.16). In the circuit of Fig. 3.16*a*, the IC switches the current I specified in the network of a high-value resistor R (the current source control mode). If the output voltage V_O from the IC exceeds the potential difference $V_0 - V_{LED}$, the diode D stays off and the current I flows through the LED that emits a stable luminous flux. As the IC changes state and its output voltage decreases, the diode D turns on and practically all of the current I switches to the input of the IC. The LED then rapidly turns off. Both transistor-transistor logic (T^2L) circuits and emitter-coupled logic (ECL) circuits can be used to advantage as input ICs.

An emitter follower T can also act as a matching element (Fig. 3.16*b*). The output current from the IC is obviously equal to 1-3% of I depending on the gain of the transistor. It should be kept in mind that the emitter follower produces current in proportion to the

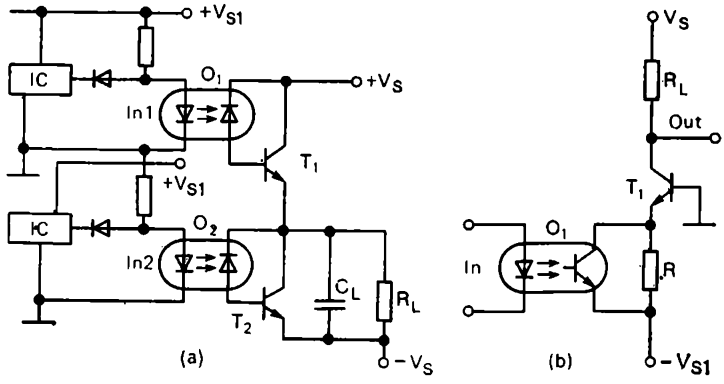


Fig. 3.17. Matching the optron output to a load
(*a*) circuit with phototransistors in series; (*b*) circuit with high-speed transistor

IC output voltage, so the current I through the LED may change on replacement of the IC by another circuit.

Consider now the conditions of efficient operation of digital optrons with subsequent stages (loads). The output current of modern diode photodetectors is relatively small (see Table 3.1); the speed of photodetectors is lower than that of electronic circuits. For this reason, optimum interfacing of the diode PD with a load calls for intermediate current amplifiers and special measures to speed up the recharge of the output capacitance of PD and load capacitance.

The circuit of Fig. 3.17*a* having two transistors in series with a pair of diode optrons can effectively drive an appreciable capacitive load. When the input *In1* is at logic 1 (high voltage level) and the input *In2* is at logic 0 (low voltage level), the optron O_1 and transistor T_1 turn on, and the load capacitor C_L charges fast as it draws a high emitter current I_{E1}. The transistor T_2 and optron O_2 remain off. With a change of input signal (logic 0 at *In1* and logic 1 at *In2*), the capacitor C_L discharges fast through the on transistor T_2.

The circuit of Fig. 3.17*b* features improved output dynamic characteristics due mainly to a decreased load resistance of the photodetector.

The network that couples the transistor photodetector to the load resistor R_L includes a high-speed transistor T_1 in a common-base (CB) configuration. The load connected to the phototransistor output is found to be a low-value input resistance of the transistor, which makes for a faster recharging of the depletion capacitance. Signal shaping at R_L is governed by the high-speed transistor.

A version of the pulse amplifier of the blocking oscillator type appears in Fig. 3.18. The diode optron here replaces a pulse transformer that is not amenable to miniaturization. The optron provides direct

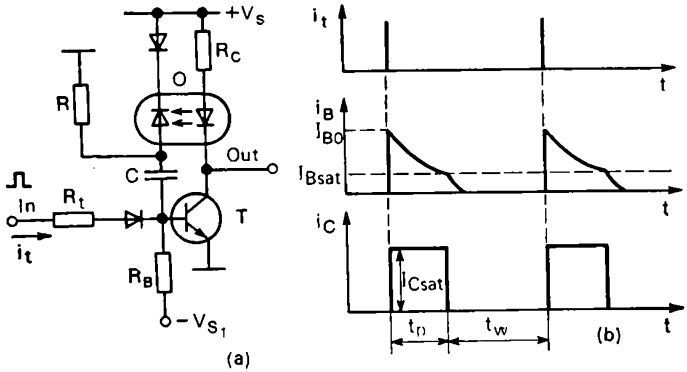


Fig. 3.18. Optoelectronic blocking oscillator
(*a*) schematic; (*b*) waveforms

current, so the circuit shapes square pulses whose width is only bounded by the time lag of the transistor and RC network parameters. What distinguishes the circuit is its high immunity with respect to the supply source.

The circuit operates in the following way. A trigger pulse i_t drives the transistor in series with the LED into its active region of operation. The transistor's collector current i_C begins to flow through the LED, and the transistor base draws the feedback current i_B via the photodiode and capacitor C. The feedback signal switches the transistor into saturation. The output voltage V_O and the collector saturation current $I_{C\,sat}$ are equal to

$$\begin{aligned} V_O &= V_S - V_{CE\,sat} - V_{LED} \\ I_{C\,sat} &= V_O/R_C \end{aligned} \tag{3.7}$$

The transistor and LED are connected in series.

Modern LEDs have a lower value of maximum permissible forward current than transistors. Therefore, it is the LED that places an upper limit on the output current, which equals 100-200 mA. The output pulse width t_p is equal to the time during which the base current drops from its initial level, I_{B0}, to the level of current at the boundary of saturation, $I_{B\,sat}$:

$$i_B\,(t_p) = I_{B\,sa} = I_{B0} \exp\,(-t_p/\tau_c) \tag{3.8}$$

where $I_{B\,sat} = I_{C\,sat}/\beta$; β is the CE current gain of the transistor;

$$I_{B0} = \frac{I_{C\,sat}}{K} \frac{R}{R + r_B}$$

K is the optron CTR; r_B is the base resistance of the transistor; and $\tau_c = (r_B + R)\,C \approx RC$.

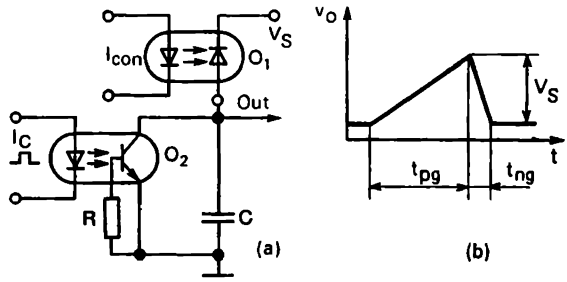


Fig. 3.19. Optoelectronic ramp generator
(*a*) schematic; (*b*) waveform

From expression (3.8) we obtain the formula for the pulse width:

$$t_p = \tau_c \ln I_{BO}/I_{B\,sat} = \tau_c \ln K\beta \qquad (3.9)$$

Thus, to shape pulses of long duration, the transistor should provide a high β because the CTR of diode optrons is less than 5%.

The space width t_w, i.e. the time between pulses, depends on the recovery time of the base circuit, which is the time taken for the timing capacitor to discharge

$$t_w = 2.3\,(R_B + R)\,C \qquad (3.10)$$

The optronic blocking oscillator, unlike the transformer-type oscillator, features enhanced thermal stability because the optron CTR decreases with rising temperature, while the β of the transistor grows with temperature. Besides, the circuit is simpler in design and more adaptable to streamlined manufacture. A limitation of the circuit is a decrease in the energy of the output signal due to a low permissible power dissipation in the LED.

Optrons used in pulse circuits can act as multifunctional elements. In other words, the pulse circuit can harness for use any properties of the optron along with the property of optical coupling. For example, by virtue of the fact that the *VI* characteristics of photodiodes exhibit regions with a large incremental resistance, diode optrons can function as dc sources to charge capacitors of ramp generators.

The schematic diagram of a ramp generator is shown in Fig. 3.19. The duration of a positive-going ramp, t_{pg}, is preset by the on state of a diode optron O_1; the time the capacitor C takes to discharge via a transistor optron O_2 specifies the duration of a negative-going ramp, t_{ng}. The output voltage peaks at about V_S. The time t_{pg} can thus be found from the expression

$$t_{pg} = V_S C/KI \qquad (3.11)$$

where K is the CTR of the optron O_1

In the voltage range 0.5-15 V, the photodiode current for modern optrons varies from 0.2 to 1 μA. Inserting a resistor R in the base circuit of the transistor of optron O_2 allows control over the trailing-edge time of the ramp pulse, which feature is taken advantage of in shaping triangular waves.

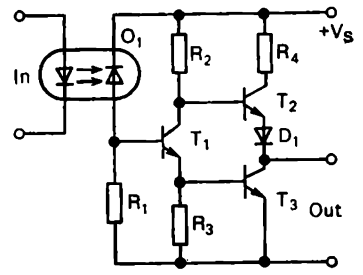


Fig. 3.20. Optoelectronic IC

Digital OICs constitute the main family of optoelectronic ICs. The efforts directed toward improving digital OIC parameters primarily aim at reducing both the switching time down to 10^{-8} s and the level of input currents, to 0.1 mA. Let us look at the principal features of a digital OIC by considering an example of the typical circuit illustrated in Fig. 3.20. This optron IC combines a diode PCP and a standard integrated-circuit amplifier mounted in a common package.

The digital OIC can switch into two static states, logic 0 state when the output voltage level goes LOW ($V_{OL} \approx 0.3$ V) and logic 1 state when the output is HIGH ($V_{OH} \approx 3.0$ V). When a nominal forward voltage flows through the LED, the diode emits and induces a photocurrent I_P in the photodiode of the optron. The photocurrent I_P drives the base of an *npn* transistor T_1 and biases T_1 on. The emitter current of T_1 arrives at the base of T_3 and causes this transistor to saturate. The output voltage thus becomes equal to the voltage drop across the saturated transistor ($V_{OL} \approx 0.3$ V). When the input current is low and the LED does not emit, only the dark current flows through the photodiode and the transistor T_1 stays off. In this case the base current I_{B2} of T_2 flows through R_2. The value of I_{B2} is such that it drives T_2 into saturation. The circuit switches to a high output level V_{OH} equal to about $V_S - V_{BET2} - V_{D1} \approx 2.5$ to 3.5 V. Thus, the output circuit of the digital OIC fully matches the circuit of the standard IC amplifier. This secures compatibility between the OIC input and the output of standard digital ICs.

The OIC switches from one logic state to the other during the transient process of switching. In this switching period the transistors operate in the active region and dissipate a large amount of power. The low-value resistor R_4 decreases the output current during the transient when both T_1 and T_2 are on. To limit the amount of power dissipated during the transient, it is necessary to preclude too slow a rise of the input current.

Digital OICs mainly apply for interfacing of various ICs and also ICs and discrete elements.

3.4.3. Analog applications of optrons. Analog (linear) optoelectronic circuits and OICs with electrically isolated input and output circuits offer a number of obvious advantages, so progress in this field

is highly welcome. However, design of these circuits at the modern stage of optoelectronics is a complex technical problem involving certain difficulties. First, as compared to electronic linear circuits, optically coupled isolators more heavily distort signals transmitted along increasingly complicated electrical channels and optical paths of limited linearity. Second, what adds to the above discomfiture is the fact that the guidelines for design of linear optronic circuits are mainly taken proceeding from the parametric methods. These methods orient the linear circuit designer to select optrons with similar parameters and characteristics. Recourse to this approach is made because of the need to introduce a certain amount of degenerative feedback unrealizable in the case of conductive decoupling with the aim to linearize the transfer characteristic in wide dynamic and temperature ranges.

Thus, for analog pulses to be transferred with the aid of an optron, it is first of all necessary to ensure the linearity of the transfer (input-output) characteristic. Only analog optrons with linear input-output characteristics can afford undistorted transfer of signals between conductively isolated circuits.

It is diode PCPs that most fully comply with the requirements imposed on analog optrons. They feature a wide range of operating frequencies, low level of intrinsic noise, and insignificant temperature dependence of parameters. These are the advantages of diode photocoupled pairs as regards their application in analog devices (Table 3.6).

Table 3.6. Comparison of the Parameters of PCPs Used in Analog Devices

PCP type	K, %	Operating frequency band	Noise level	Dark current	Temperature dependence of parameters
Diode	0.5-2	Very wide	Very low	Very small	Weak
Transistor	50-100	Narrow	Moderate	Moderate	Significant
Resistor	30-100	Very narrow	Low	Very small	Very strong
Diode-transistor	10-40	Wide	Moderate	Small	Significant

The current transfer ratio K for a diode PCP is easy to increase by placing an IC amplifier into its output circuit. What presents the main difficulty in designing is primarily a narrow region of linearity of the transfer characteristic. A typical I_O-I_I relation for the diode PCP (Fig. 3.21) is linear in the range from $I_{I\,min}$ to $I_{I\,max}$. At low input currents when $I_I < I_{I\,min}$, the emissive power of the LED does not yet reach its nominal level. At high input currents when $I_I > I_{I\,max}$, the structure begins to heat up, causing the temperature drift of parameters. Fig. 3.21 also illustrates the plot of the normalized current transfer ratio K/K_0 versus I_I, where $K_0 = I_O/I_I$ is the CTR in the linear section of the transfer characteristic. It

is seen that the CTR drops off at the edges of the operating region (when $I_I = I_{I\,min}$ and $I_I = I_{I\,max}$). In the specified range of changes of the input current, the CTR nonlinearity is given in percent as

$$\frac{K_0 - K}{K_0} = \frac{\Delta K}{K_0} 100 \qquad (3.12)$$

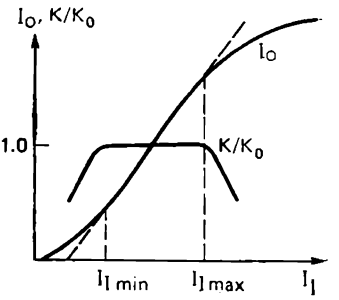


Fig. 3.21. Transfer characteristic of a diode PCP

Provided the thermal stability is good, modern diode PCPs can transfer analog signals with a linearity to better than 2 % at $I_{I\,max}/I_{I\,min} \approx 2$ or 3. A decrease in nonlinearity sharply narrows the working range of currents. What complicates the task of undistorted analog pulse transfer through a PCP is that the parameters heavily vary with operating conditions and experience drift with time.

The linear differential amplifier illustrated in Fig. 3.22 may serve as a typical example of the optoelectronic analog circuit. The pair of transistor PCPs forming an optical channel for signal transfer must be perfectly symmetric and exhibit identical parameters and characteristics. In the quiescent mode, the same currents flow through the LEDs of both PCPs, $I_{LED1} = I_{LED2}$. The photocurrents I_{P1} and I_{P2} now come in the opposite directions and have no effect on the output amplifier A_3. Possible time- and temperature-dependent changes in the static conditions and parameters are mutually offset. As an input signal V_I arrives at the stage A_1, this stage transmits the signal to the photocoupled pair O_1, which then transfers it to A_3 for further amplification. This device is unable to compensate for the nonlinearity of the transfer characteristic of O_1. As a result, at a large value of V_I the transfer ratio K_1 of O_1 changes, while K_2 of O_2 remains invariable since I_{LED2} is kept constant. The amplification linearity is accurate to 1-5 % and the CTR of the device stressed for 100 000 hours at 25°C degrades by about 80-95 %.

The use of differential optrons appreciably improves the quality of analog signal transfer. To illustrate how a differential optron improves the linearity of the transfer characteristic, consider an example of the circuit of Fig. 3.22*b*. The LED illuminates two single-type photodiodes D_1 and D_2 having identical parameters. The current I_{LED} is a function of the feedback current I_{P1} as well as input current I_I:

$$I_{LED} = K_1 (I_I - I_{P1}) = K_1 (I_I - K_{D1} I_{LED}) \qquad (3.13)$$

where K_1 is the gain of the amplifier A_1; K_{D1} is the CTR of the LED-D_1 pair. From (3.13) we have

$$I_{LED} = K_1 I_I / 1 + K_1 K_{D1} \qquad (3.14)$$

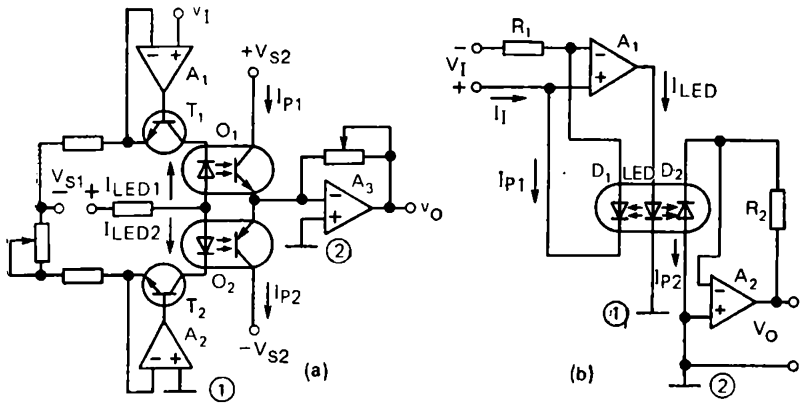


Fig. 3.22. Analog optoelectronic ICs
(*a*) parametric differential amplifier; (*b*) with differential optron

When the amount of feedback is large, K_1K_{D1} is much higher than 1, the current through the LED is $I_{LED} = I_I/K_{D1}$ and the photocurrent of the LED-D_2 pair is $I_{P2} = K_{D2}I_I/K_{D1}$. The parameters K_{D1} and K_{D2} for the single-type pair are the same and vary to the same degree, with the result that i_{P2} (t) and i_I (t) are equal in magnitude and independent of the nonlinearity and instability of optron characteristics. The device using this differential optron amplifies linearly to within 0.01-0.2%; the CTR degradation in 100 000 hours is about 99%.

Of much interest is the application of linear optrons in measurement technology. Traditional methods for measuring voltages above 5 kV require (1) high-voltage isolation transformers, (2) ungrounded synchronized oscillographs, or (3) voltage dividers to measure the difference of voltages with respect to ground. These methods are laborous, call for cumbersome equipment, and offer low accuracy. The use of analog optrons as measuring devices holds good promise for decreasing the size and mass of equipment and improving the measurement accuracy of high voltages. The main advantage of such devices is that they are fast and ensure good isolation in a simple way. Fig. 3.23 illustrates examples of the circuits of photoelectronic meters. In the circuit of Fig. 3.23*a*, the LED is placed in series with a resistor shunting a power device to increase power consumption and to avoid measuring reverse voltages. In the circuit of Fig. 3.23*b*, the signal V_{div} to be measured arrives at a LED from a step-down voltage divider. The LED operates with a constant reference bias V_{ref}. Next the light signal emitted from the high-voltage LED and modulated in voltage V_{div} comes to a photodetector PD and then to a measuring device such as an electronic oscillograph (EO). The advantages of such a device are: (1) small size and mass, thus

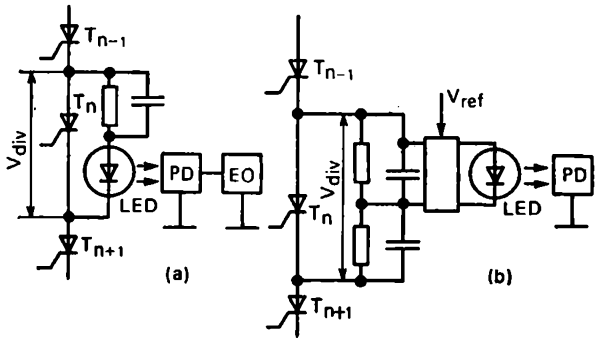


Fig. 3.23. Optoelectronic transducers

allowing the device to be built into test equipment; (2) high input resistance; (3) possibility of servicing grounded equipment on the side of measurements.

3.4.4. Optron switches. These switches connected to weak-current control devices can operate the elements of high-current high-voltage circuits. For example, power thyristor PCPs controlled by integrated circuits are able to switch voltages up to 2 000 V and over and currents up to 500 A and over. These devices serve the same purposes as ordinary thyristors. Thyristor PCPs of lower power or photodarlington PCPs act as switches in pulse shapers of moderate power, with a pulse current level of 1 to 5 A. The circuit of such a pulse shaper designed to trigger a power thyristor is shown in Fig. 3.24. The capacitor C preliminarily charges from the supply source of a power thyristor Y via a resistor R and diode D. The capacitor then discharges via an optoisolator thyristor, control resistor R_{con}, and control junction of the thyristor Y as the control unit CU delivers an input pulse to a thyristor optoisolator O_1. The optoisolator O_1 switches off fast since by the moment of switch-off, a weak current $\Delta V/R$ flows through it, where ΔV is the difference of voltage drops across Y and O_1.

3.4.5. Application of special optrons. Optrons using air as an open optical path offer unique potential in various applications. A chopped-beam-responsive optron, for example, discriminates an object moving through its air gap; a reflected-beam responsive optron identifies an object or its state as it picks up the LED beam reflected from the object surface (see Fig. 3.11). The CTR of these optoisolators is small, so they use IC amplifiers to shape signals of the desired amplitude.

An example of the block diagram of the air-gap optron as a sensor of an object edge appears in Fig. 3.25. The device consists of the transistor optron proper, amplifiers A_1 and A_2, pulse generator G, detecting flip-flop F/F, and LED display. As the edge of an object

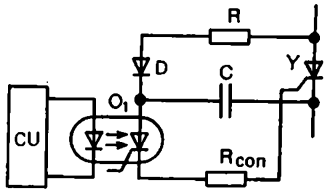


Fig. 3.24. Optoelectronic pulse shaper

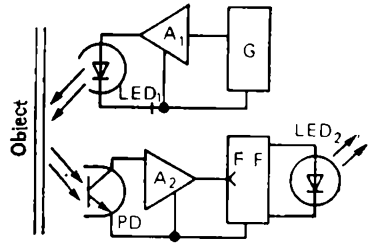


Fig. 3.25. Air-gap optron as sensor of an object edge

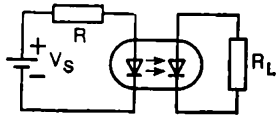


Fig. 3.26. Diode PCP operating in the photovoltaic mode

reflects the LED_1 beam and enables optical coupling between the LED_1 and PD, the signal from PD triggers the F/F into the opposite state and the LED_2 begins to emit.

Air-gap optrons can also be used for detailed and accurate investigation of objects, in particular, analysis of surface defects, detection of micromotions, etc.

Figure 3.26 illustrates a diode PCP operated in the PV mode and used as an energy source. The main requirement on these PCPs is that they must convert energy at a high efficiency. The Q factor for this energy source is given by

$$Q_B = P_O/P_I \tag{3.15}$$

where P_O is the maximum power taken off PD into the load. For a typical VI characteristic of the PCP, $P_O = aI_{sc}V_{oc}$, where $a = 0.7$ to 0.8, I_{sc} is the short-circuit current, and V_{oc} is the open-circuit voltage. Hence,

$$Q_B = aKV_{oc}/V_I \tag{3.16}$$

where K is the CTR for a short-circuited output. It is clear from (3.16) that these PCPs should preferably use PDs based on wide-bandgap semiconductors which ensure a high value of V_{oc}.

The PCPs used as supply sources offer all the advantages of optical coupling. For example, in the transducer illustrated in Fig. 3.27*a*, a noise signal appearing in the power unit PU quite readily travels to the control unit CU along the common supply line and disturbs the CU operation. In the transducer of Fig. 3.27*b* where a DT PCP

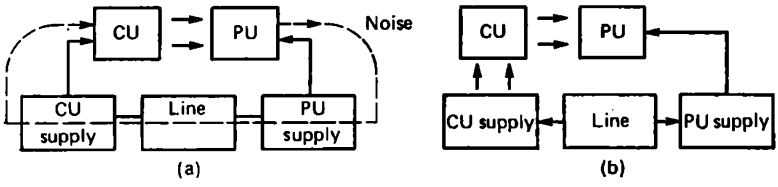


Fig. 3.27. Illustrating improvement in noise immunity with the aid of an optron
(*a*) scheme showing passage of noiso signal along supply line; (*b*) control unit (CU) supply source that electrically isolates CU from power unit

serves as a supply source for CU, the above passage for a noise signal is shut down since the optron totally isolates the control circuit from the power unit.

3.5. Reliability of Optoelectronic Devices

3.5.1. Reliability parameters. A number of factors affect the operating lifetime of optoelectronic semiconductor devices. The failure rate depends on the time to failure, operating conditions, and ambient temperature. Devices may fail under the action of shocks, vibrations, acceleration, and acoustic noise, and also on exposure to low or elevated pressures, enhanced humidity, and other external influences. In a number of applications, devices must be exceptionally resistant to solar, electron, proton, gamma, X-ray, and neutron types of radiation.

Experience in industrial production and operation of photoelectronic devices reveals that failures are probabilistic in character, and the relationships for the statistic description of failure processes are the same regardless of the type of semiconductor device. In a first approximation, the failure rate (mean failure density) can be taken to obey an exponential law. The reliability parameters include:

— time-to-failure density function

$$f(t) = \lambda_0 \exp(-\lambda_0 t)$$

— probability of survival function (reliability function)

$$P(t) = \int_t^\infty f(t)\,dt = \exp(-\lambda_0 t)$$

— mean life

$$t_m = 1/\lambda_0$$

where λ_0 is the failure rate.

In addition to these parameters, manuals on optoelectronic devices also cite the minimum up time, which is the minimum time period of operation of a device in specified conditions without malfunctions.

It should be remembered that specifications and manuals list peak values of λ_0 and t_m obtained in tests under aggravated conditions. Practice shows that the failure rate observed in actual operating conditions is approximately one tenth that estimated in production tests.

3.5.2. Physical causes of failure. In assessing the failure probability of devices, one distinguishes between two types of failure, creeping (incipient, or parametric) and complete (catastrophic). The first type is indicative of a partial failure that appears as a deviation of at least one of the parameters from the rating. The parameters which specify a creeping failure are known as reliability (validity) criteria of a device. The second type of failure shows up as complete inability of a device to perform its function due to breaks, shorts, and other malfunctions.

The results of studies on optoelectronic devices subjected to prolonged tests allow us to single out the following physical causes of failure:

— degradation of a photoemitter with an attendant decrease in its external quantum efficiency and CTR (failure rate λ_{em});

— degradation of a photodetector that impairs the parameters of its output circuit λ_{PD1} and also other parameters on exposure to, say, a strong electric field built up between the light source and PD;

— deterioration of the structure of a device due to breaks of internal wire connections, cloudiness or separation of the immersion medium, current leakage between the input and output circuits over the inner or outer surface of the package, etc. (λ_h).

Assuming that failure mechanisms act independent of one another and the exponential law holds for each mechanism, the total failure rate λ_Σ can be written as

$$\lambda_\Sigma = \lambda_{em} + \lambda_{PD1} + \lambda_{PD2} + \lambda_h$$

The failures due to incorrect application of devices are not taken into account. The values of all failure rates are not calculated theoretically, but estimated empirically based on numerous tests for durability (service life). For example, empirical results firmly confirm that the photoemitter's external quantum efficiency degrades faster with an increase of both the operating current through the LED and ambient temperature. First, the above factors reduce the injection ratio because of the changed structure of the space charge near the *pn* junction and the growing role played by nonradiative recombination in the junction region and on the surface. Second, these factors reduce the LED's internal quantum efficiency through manu-

facturing defects such as dislocations, microcracks, surface states, chemical contaminants, and impurities doped in excessive concentrations.

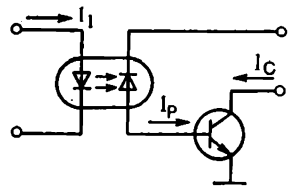


Fig. 3.28. Diode-transistor optron

3.5.3. Reliability criteria. To these criteria commonly belong the current transfer ratio K, leakage current I_{lk} at the output of a *pn* junction photodetector, isolation resistance $R_{I/O}$, and input voltage V_I. These parameters are chosen as reliability criteria because they describe the input-output, transfer, and isolation properties of an optron. The stability of the above parameters over the specified period with a minimum of deviation from the ratings is testimony of the optron serviceability.

As mentioned earlier, the main cause responsible for the CTR degradation is the degradation of photoemitter properties. Besides, the transmittance of an optical channel, T_{OC}, photodetector responsivity S_P, and the gain B of the output amplifier also degrade with time. For a diode-transistor PCP displayed in Fig. 3.28, the CTR is equal to

$$K = I_C/I_I = T_{OC}S_P\eta\ (I_I,\ t)\ B\ (I_P) \tag{3.17}$$

where I_C is the collector (output) current; I_I is the input current through the LED; η is the LED quantum efficiency; and I_P is the photocurrent.

The parameters S_P, T_{OC}, and I_I are not practically time-dependent. The gain B varies as a function of photocurrent which, in turn, depends on the degree of degradation of the LED emissive power. So, the gain B also varies with time in proportion to the LED degradation. From expression (3.17), we can then relate the change in K over time to the changes in the LED efficiency and gain:

$$\frac{\Delta K}{K} = \frac{\Delta\eta}{\eta}\left[1 + \left(1 + \frac{\eta}{\Delta\eta}\right)\frac{\Delta B}{B}\right] \tag{3.18}$$

where $\Delta B = B\ (I_{P\,f}) - B\ (I_{P\,in})$; $I_{P\,f}$ and $I_{P\,in}$ are the final and initial photocurrents respectively; and $\Delta\eta/\eta$ is the normalized change in the LED efficiency for a specified operating time.

For example, suppose a PCP operated at 16 mA I_I and the LED efficiency degraded in 10 000 hours of service by 15% so that $\Delta\eta/\eta = 0.15$. From the I_P-I_I relation we can now determine the initial photocurrent $I_{P\,in}$. Let it be equal to 24 μA and, correspondingly, $B\ (I_{P\,in}) = 145$. A change in photocurrent I_P with the LED degradation is given by

$$I_P = \left(1 + \frac{\Delta\eta}{\eta}\right) I_{P\,in} \tag{3.19}$$

whence we find that I_{Pf} is 20 μA, which corresponds to a B of 150. Thus, $\Delta B/B = 5/145$ so that $\Delta K/K = -0.12$.

In general, $|\Delta K/K| < |\Delta\eta/\eta|$ if $\Delta B/B > 0$ and $|\Delta K/K| > |\Delta\eta/\eta|$ if $\Delta B/B < 0$. However, if $\Delta B/B$ is small, the rate of CTR degradation will be approximately equal to photoemitter degradation.

The output amplifier added to the PCP enables the LED to operate at lower currents. This allows reducing the rate of emitter degradation and, hence, the rate of optron degradation. Besides, $\Delta B/B > 0$ in this range of currents, which, as mentioned above, tends to decrease the rate of optron degradation with respect to that of emitter degradation.

The parameters of optrons are liable to changes due to the effect of a high isolation voltage $V_{I/O}$. If the output circuit is biased positive with respect to the input circuit, the migration of ions of uncontrollable alkali metal impurities produces a negative charge on the surface of the PD crystal; the resultant conductive paths give rise to leakage current I_{lk} in the output circuit. With the negative bias on the output circuit, the positive charge induced on the crystal surface deforms the space charge region so heavily that the optron CTR diminishes. The leakage current markedly grows with the isolation voltage and ambient temperature.

3.5.4. Radiation stability. The parameters of optoelectronic devices are very sensitive to the effect of ionizing radiations. For example, neutrons or γ-rays incident on a GaAs LED produce crystal defects which become nonradiative recombination centers reducing the light output from the LED. It is found that the concentration of radiation-induced defects linearly rises with radiant intensity.

The theoretical analysis gives the following expression to describe a decrease in the emissive power of a LED on exposure to ionizing radiation:

$$P = P_0 (1 + \tau_0 k_{\gamma, n} \Phi_{\gamma, n})^{-m} \qquad (3.20)$$

where P and P_0 are radiated powers of the LED before and after irradiation; τ_0 is the carrier lifetime before exposure to ionizing radiation; $k_{\gamma,n}$ is the proportionality factor accounting for the kind of radiation; and m is the exponent of the power function, which accounts for voltage on the pn junction of the LED.

From expression (3.20) it is readily seen that as the values of τ_0 and $k_{\gamma,n}$ decrease, the resistance of a photoemitter to ionizing radiation grows. Silicon-doped GaAs photoemitters obtained by liquid-phase epitaxy are most stable to radiation. The radiation stability of a GaAs emitter substantially increases if aluminum atoms substitute for a fraction of gallium atoms in the crystal lattice. An addition of merely 1 % Al increases the γ-radiation stability more than twofold.

Ionizing radiations change the dynamic parameters of optoelectronic devices as well as irreversibly reduce the CTR. The investigations of the effect of radiations on optoelectronic circuits have revealed that a γ-radiation exposure dose of 5×10^4 gigaroentgens ($5 \times \times 10^6$ rads) increases the turn-on delay time two- or threefold.

The above mentioned changes in the properties of devices occur in the time period after exposure to radiation. During irradiation the parameters may change so heavily that the device will fail to operate at all. The degraded-performance recovery time after the effect of radiation, i.e. the time it takes to acquire a new equilibrium state, is close to 3×10^{-7} s for GaAs(Si) emitters and to 3×10^{-8} s for GaAs(Zn) emitters.

The data presented in this section show that modern optoelectronic devices can in principle withstand the neutron flux whose density approaches 10^{12}-10^{13} cm^{-2} and γ-radiation doses of 10^5 GR.

Chapter 4

Coherent Optoelectronics

4.1. Physical Processes in Lasers

4.1.1. General. As mentioned in the introduction to this book, coherent (laser) optoelectronics opens qualitatively new ways for the development of electronics and offers such basic advantages as increased information capacity of optical paths, space-time modulation of optical signals, generation of narrow beams, and enhanced recording density. More rapid realization of the great potential of coherent optoelectronics depends on the progress in the area of its building blocks. At present, lasers, optical modulators, and light guides have stepped over the threshold of experimental development and are now in commercial production, while holographic memories, integrated optic devices, and space-time modulators are still at the stage of laboratory investigations.

Nevertheless, even today there are indications that permit us to speak of a great future of coherent optoelectronics. Specialists believe, for example, that the time is not far distant when in the main only two types of communication will be in general use, namely, radio communication if at least one of the distant stations is in motion and coherent fiber-optics communication to establish contact in all other cases.

There are a few trends in coherent optoelectronics toward more and more extensive use of lasing emitters in data processing systems.

In this connection, scientists pin strong hopes on optical (mainly holographic) memories and integrated optic devices. The domain of optical memories makes the best use of such attractive features of coherent emission as low beam divergence and spatial modulation. The domain of integrated optics primarily tends to extend the principle of manufacturing integration to cover the elements of laser optoelectronics.

Coherence is the main distinctive feature of all devices of this branch of optoelectronics, first of all, coherent sources—lasers (see Subsec. 1.1.3).

Assume for simplicity that an individual atom emits radiation in the form of a plane monochromatic wave. This wave propagates in space and time according to the expression known from physical optics:

$$E(x,\ t) = E_m \cos\left[(2\pi\nu t - 2\pi\nu n/c_0)\, x + \varphi_0\right] \qquad (4.1)$$

where E is the electric field strength of the wave; E_m is the field strength swing; ν is the oscillation frequency; n is the refractive index of a medium; c_0 is the light velocity in a vacuum; x is the wave propagation direction; and φ_0 is the initial phase of the wave.

Actual optical radiation of a physical body represents a superposition of electromagnetic waves emitted by a large number of excited atoms. If each atom emits independent of others so that the values of parameters ν and φ_0 and also the directions of polarization are different for all atoms, then incoherent radiation takes place.

If all atoms emit consistently (coherently) in time so that the values of ν and φ_0 and the directions of polarization for all atoms are identical, in evidence is coherent emission of the resultant radiant flux described by expression (4.1).

At present, lasers are the only sources of coherent radiation, which generate concentrated beams of strictly fixed oscillation frequency and high monochromaticity.

4.1.2. Light amplification. As its name implies, which is the acronym of light amplification by stimulated emission of radiation, the laser uses stimulated (induced) emission of radiation to generate light coherent in space and time. To put it briefly in a very simplified way, the process of stimulated emission arises in the following way. An external photon incident on an excited atom causes the atom to pass to another energy state and to emit a photon whose energy is equal to the stimulating photon. If we have a system of active atoms (active medium) and stimulate the system by electromagnetic radiation, light amplification is likely to occur if the number of photons generated by the induced radiation exceeds the radiation loss by absorption and scattering.

Consider the process of laser amplification in more detail. The energy of external excitation (pumping), E_p, preliminarily promotes

a fraction of particles from low energy levels E_1 to higher levels and further to excitation levels E_2 (Fig. 4.1.). As they return from level E_2 to level E_1 (eV), the particles emit photons of wavelength (μm)

$$\lambda = 1.23/(E_2 - E_1) \quad (4.2)$$

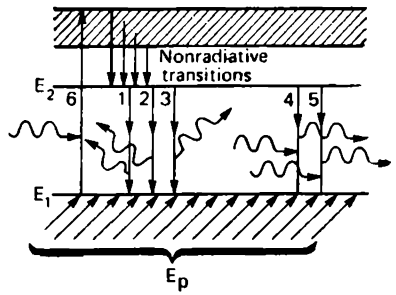


Fig. 4.1. Quantum transitions in laser material

The process in which particles decay from level E_2 and return to level E_1 may occur in different ways. Spontaneous transitions are possible, where particles emit independent of one another in time and give up photons with different directions of polarization, so that the resultant flux can only be defined by statistical average parameters (transitions *1*, *2*, and *3* in Fig. 4.1). This process of motion of the emitting particles from the excited to the equilibrium state does not involve stimulating photons and only gives rise to incoherent radiation.

Concurrent with spontaneous transitions, induced transitions from energy state E_2 to state E_1 are possible (transitions *4* and *5* in Fig. 4.1.). Such transitions arise from the action of stimulating photons. All active particles emit almost simultaneously in the interrelated manner, so that the emitted photons and stimulating photons are all alike in properties. This process produces a coherent flux. So, stimulated emission is coherent emission due to induced transitions, where photons have the frequency, phase, polarization, and direction of propagation identical to those of original exciting photons.

Determine the conditions of amplification of stimulated emission. The energy levels used in light amplification or lasing are called laser levels. Correspondingly, the induced transition between laser levels or energy bands is a laser transition of wavelength λ defined by expression (4.2). Along with laser transitions *4*, *5* and spontaneous transitions *1*, *2*, *3*, particles make transitions from level E_1 to a higher energy state than E_2, thus leading to absorption of exciting radiation (transition *6* in Fig. 4.1.).

Laser amplification can take place only if the number of laser transitions is greater than that of spontaneous transitions and absorptive transitions. The number of laser transitions in time Δt can, to a first approximation, be expressed as

$$n_l = B_{21} Q_{ex} N_2 \Delta t \quad (4.3)$$

where B_{21} is the probability of a laser transition; Q_{ex} is the energy of exciting radiation; and N_2 is the concentration of atoms excited to energy state E_2.

Spontaneous transitions from E_2 to E_1 do not depend on exciting radiation and make no contribution to coherent emission. In a first approximation, the number of these transitions can be written in the form

$$n_{sp} = -A_{21}N_2\Delta t \tag{4.4}$$

where A_{21} is the probability of a spontaneous transition from E_2 to E_1.

The number of quantum transitions resulting in the absorption of exciting radiation is given by

$$n_{abs} = -B_{12}Q_{ex}N_1\,\Delta t \tag{4.5}$$

where B_{12} is the probability of an absorptive transition.

Assuming to a first approximation that probabilities B_{21}, B_{12} and B are equal to each other, the condition of laser amplification takes the form

$$B\,(N_2 - N_1)\,Q_{ex} - A_{21}N_2 > 0 \tag{4.6}$$

At a low level of spontaneous radiation, the required condition of laser amplification is defined as

$$BQ_{ex}\,(N_2 - N_1) > 0 \quad \text{or} \quad \Delta N = (N_2 - N_1) > 0 \tag{4.7}$$

In the equilibrium state of the system, N_2 is always lower than N_1, and laser can amplify light only under the action of external influences such as charge carrier injection, gas discharge, and optical or electron-beam excitation (pumping).

Thus, laser amplification derives from the fact that exciting radiation acquires more energy than it loses by absorption as it propagates through the laser material.

Obviously, the higher the probability of a laser transition, B_{21}, the more effective is the laser amplification. Laser amplification improves as the probability of laser transitions and the density of energy states in bands become higher. Solid-state and gas lasers rely on transitions between discrete levels in ions, atoms, or molecules. They amplify more weakly than semiconductor lasers and thus are much larger in size.

4.1.3. Population and population inversion. The notion of energy level population is used for quantitative estimation of laser amplification. By energy level population is meant the number of atoms in a unit volume, which occupy the same energy levels. In the conditions of thermodynamic equilibrium, the population of energy levels obeys the Maxwell-Boltzmann statistics:

$$N_2/N_1 = \exp\left[-(E_2 - E_1)/kT\right] \tag{4.8}$$

where N_2 is the population of atoms excited to state E_2; and N_1 is the population of atoms remaining in state E_1. The difference here

$$\Delta N = N_2 - N_1 = N_1 \left[\exp\left(-\frac{E_2 - E_1}{kT} \right) - 1 \right] \qquad (4.9)$$

is negative, and the material has a normal population when N_2 is lower than N_1. This is the equilibrium condition for the system where laser amplification cannot occur.

When $\Delta N > 0$ secured by pump energy, the condition of inverse population sets in, and the transmitted radiation can be amplified at the expense of energy of excited atoms.

From expression (4.9) it formally follows that the condition $\Delta N > 0$ holds at absolute negative temperature ($T < 0$). Therefore, the condition of population inversion is sometimes referred to as a negative temperature condition. A medium where the inversion of population occurs is known as an active medium.

Laser amplification thus calls for population inversion ($N_2 > N_1$) and suppression of spontaneous emission (light noise). The lowest level of pump energy which satisfies the inversion condition is termed inversion threshold.

4.1.4. Lasing. For the above described laser amplifier to be able to lase, a positive feedback (PFB) loop must be built around it. The parameters of the PFB loop are so chosen that the radiation energy fed back from the output to the input of the laser amplifier should be enough to compensate for the energy lost in the closed PFB circuit.

Optical resonators commonly perform the function of a PFB link. The simplest resonator consists of two mirrors to enable multiple transmission of the radiated wave through the active medium; the mirrors are made semitransparent to effect light extraction. In the general case, the resonant cavity is a system of reflecting, refracting, focusing, and other optical elements enclosing an active medium to enable excitation of optical waves. The avalanche process of lasing that causes photon avalanching in the optical resonator is illustrated in Fig. 4.2*.

Thus, with PFB secured and laser amplification conditions fulfilled, all prerequisites exist for laser self-excitation. The process of excitation of a laser material that culminates in the buildup of an active medium is known as laser pumping. The pump energy is much higher than the energy of laser emission (the laser efficiency is only a few percent). In other words, the laser is an uneconomical

* The photon avalanche is similar to mudflow: fairly lengthy thawing of snow is reminiscent of pumping, and fast release of accumulated water is suggestive of a photon avalanche; but the strong flow of sill weakens fast without 'feedback'.

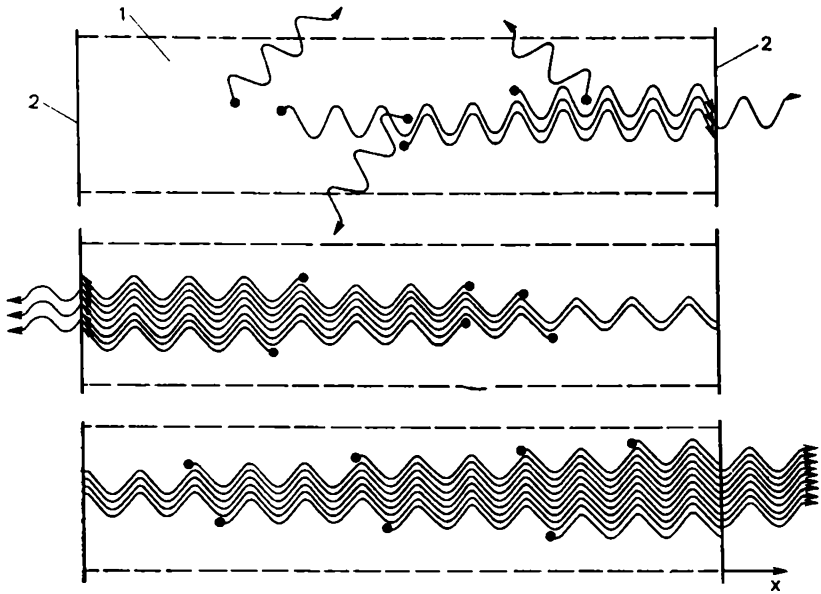


Fig. 4.2. Photon avalanching in an optical resonator (*1*) active medium; (*2*) mirrors

generator, but the laser energy is unique as regards its coherence and density in space and time.

So, the laser as a generator of optical waves must contain a pumping system and a material where an active medium can be built up in the course of pumping. There are a few varieties of laser pumping depending on the form of pump energy.

1. Optical pumping by the radiant energy of a lamp, LED, or another laser.
2. Electric-field pumping, which is electric energy excitation, specifically of injection lasers.
3. Electron-beam pumping.
4. Chemical pumping due to reactions in a laser material.

4.1.5. Lasing threshold. The laser beam is not just a flux of energy such as a solar beam, but a coherent, exceptionally ordered, pointed flux concentrated within a small solid angle. However, for these exceptional properties imparted to the laser beam we have to pay dearly — the laser efficiency is merely a few percent; tens of joules of pump energy need be spent for each joule of laser emission. The energy density of laser emission is huge; even more than that attainable in a nuclear explosion, in the order of 10^{10} J/cm^3.

An avalanche-like growth of laser energy in the active medium along the x axis of the resonant cavity of Fig. 4.2 well obeys the

exponential equation

$$E_x = E_0 \exp[(K_l - K_Q)x] \tag{4.10}$$

where E_x is the radiated energy along the x axis; K_l is the linear laser gain (on-x axis laser gain), whose value is proportional to the pump energy E_p; and K_Q is the radiation loss factor for the resonator and active medium.

For a simple linear resonator, this loss factor assumes the form

$$K_Q = \frac{1}{\chi_0} + \frac{1}{l_{res}} \ln \frac{1}{\sqrt{R_1 R_2}} \tag{4.11}$$

where $1/\chi_0$ is the absorptance of the active medium; l_{res} is the resonator length; R_1 and R_2 are the reflection factors for resonator mirrors.

The second term in expression (4.11) accounts for radiation edge loss referred to unit length of the radiator.

At a certain value of pump energy, E_{TH}, called lasing threshold, the gain K_l becomes higher than the loss K_Q, which implies an avalanche amplification of laser emission, i.e. initiation of the lasing action. The lasing threshold E_{TH} is that amount of pump energy E_p or power P_p at which the laser gain at the laser-emission frequency is equal to the energy lost in the resonator at the same frequency.

The laser radiation pattern specifies the angular spread of the beam (beam divergence), which depends on the ratio of the emission wavelength to the resonator length. The beam divergence is thus written as

$$\theta = \sqrt{\lambda / l_{res}} \tag{4.12}$$

Let us note in conclusion that in any resonator the condition of resonance exists for many types of oscillation, called modes, differing from each other in frequency and electromagnetic field distribution. The spectrum of laser radiation consists of a set of modes. Adjustable mode filters are used to extract single-mode radiation.

4.2. Laser Structure, Parameters, and Modes of Operation

4.2.1. Laser structure. Fig. 4.3 displays the block diagram of a laser which contains the following basic components:

(1) a lasing emitter, which is the principal functional unit designed to convert the pump energy to laser radiation;

(2) an external power source to provide pump energy well above threshold.

The lasing emitter, in turn, consists of a number of structural components, among which we can single out the following:

(1) an active element (elements) which is the laser proper containing a material excited under pumping to provide an active medium;

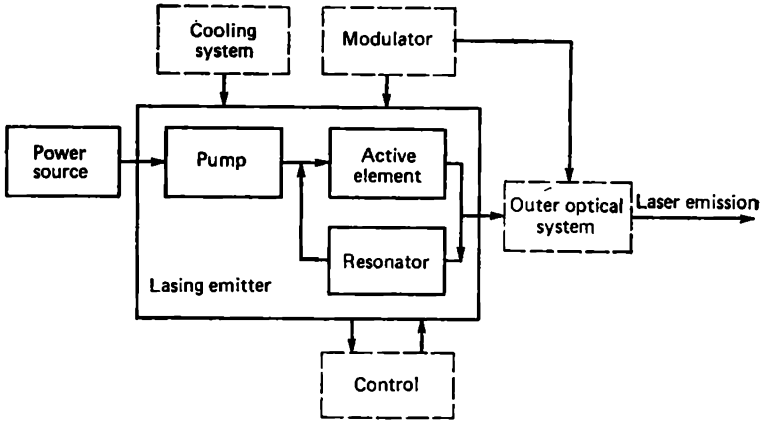


Fig. 4.3. Block diagram of a laser

(2) a pumping system comprising a set of elements intended to convert the supply energy and transfer it to the active element;

(3) an optical resonator providing PFB.

In semiconductor lasers, the emitter is structurally a single unit; in gas and solid-state devices, the emitter consists of discrete, separable elements.

The block diagram of a laser may include a number of additional units intended to ensure proper performance or control over laser emission. These can be a cooling system for the active element, modulator, external optical medium, control unit to maintain the specified values of radiation parameters, and others.

Depending on its application, a laser may or may not contain all additional units. As will be clear from the discussion of laser types, the temperature of an active material plays an important part in obtaining population inversion. In some active materials this inversion can be set up only in the conditions of deep cooling. The cooling system maintains the temperature of the active material and the pump at the required level.

Light modulation enables laser emission control and is essential for lasers employed in data processing systems. Modulation can be accomplished by changing the mode of operation of the pump or active element and also by transforming the output laser emission.

The external optical system serves to shape the laser beam to a desired pattern, for example, to change the angle of divergence. The control unit can be any device adapted to measure the radiant power, frequency, etc. and to maintain these parameters at the specified values.

Figure 4.3 displays a relatively complete block diagram of a laser. Here, full-line boxes represent basic components and boxes shown by dash lines represent auxiliary components whose number may vary with the laser type and application.

By the type of active material used, lasers fall into four basic classes: (1) semiconductor lasers; (2) gas lasers; (3) solid-state lasers; and (4) liquid lasers using liquids as active media.

4.2.2. Basic parameters of lasers. The set of spatial-temporal parameters include:

(1) laser (emission) frequency ν, which is the central frequency (or central wavelength) of the emission spectrum;

(2) laser linewidth $\delta\nu$, which is the distance between the points on the spectral line at half the peak line strength;

(3) laser beam divergence θ, specified by a plane or solid angle characterizing the angular distribution of laser energy or power;

(4) settling (readiness) time t_{set}, which is the time required for the laser to attain the nominal values of its parameters from the moment of switch-on.

The energy parameters are energy E and power P of laser emission, the first determines the energy capability of a laser and the second specifies the energy density in time. This seemingly obvious distinction between the two parameters of a laser requires a certain elucidation.

The remark that the power of a laser is equal to the power of such and such hydroelectric station became a commonplace. But then one should remember that emission of this power keeps on for merely 1 ns. The density of radiant energy (power) emitted by a laser specifies the energy (power) per unit cross-section of the beam.

One more energy parameter of a laser is its optical output efficiency η_0, defined as the ratio of the mean radiant energy (power) to the mean pump energy (power). The threshold E_{TH} or P_{TH} also characterizes the energy capability of a laser.

4.2.3. Modes of operation. A laser can operate in three modes referred to as: (1) continuous wave (CW) lasing; (2) pulse lasing; and (3) repetitive-pulse lasing (spiking).

The power of a CW laser emitting at the central frequency does not go to zero on the time interval considerably in excess of the pulse period, i.e. the laser emits continuously for a long period.

A pulsed laser emits extremely short pulses, fractions of a nanosecond in duration. Even if the pulse energy is small, the pulse power related to this negligible period of lasing proves enormous. Modern pulsed lasers, mainly solid-state types, can give outputs up to 10^8 kW in a 1-J pulse up to 0.01 ns long.

In the spiking mode of operation, lasers periodically emit trains of submicrosecond pulses.

4.3. Semiconductor Lasers

Semiconductor lasers most easily interface with microelectronic devices. The first semiconductor lasers appeared in 1962. At present two types of semiconductor laser have extensive uses. Of these, most popular is the injection laser which operates on the principle of injection of minority carriers through the *pn* junction into degenerate regions of the semiconductor to achieve population inversion in the active medium. The other type that holds much promise is an electron-beam-pumped laser which gives a coherent beam in the course of its excitation by the flux of high-energy electrons. Injection lasers use luminescent direct-gap semiconductors where the exchange of carriers between the conduction band valley and valence band top occurs with a high probability, i.e. the materials in which radiative recombination of free holes and electrons, as they execute band-to-band transitions, proceeds intensively.

4.3.1. Laser amplification in semiconductors. The properties of semiconductor lasers certainly depend on the structures of semiconductors used. Semiconductor atoms lie at quite definite, stationary, energy levels (states). The atoms accommodated in a certain energy state have a definite configuration of the electron shells. In a semiconductor, the outer (valence) shells come in contact with, or even overlap, one another. Valence electrons become free, as it were, from their shells: the electrons of one atom in a certain energy state can jump to a neighbor atom in the corresponding energy state without losing their energy, so they can travel inside the semiconductor. Individual atomic orbits thus combine to form shared (collective) orbits, with the subshells of individual atoms merged into bands. As a result of this process, discrete electronic levels of isolated atoms split into relatively wide allowed energy bands separated by forbidden bands (bandgaps). Each allowed band consists of a large number of levels equal to the number of atoms in the semiconductor. With a bandgap energy of 1 eV, the spacing between neighbor levels in the allowed band is approximately 10^{-22} eV. The bands can be regarded to be continuous.

The feature specific to emission in semiconductors is that the process of interaction involves two bands with a finite number of levels rather than merely two levels, namely, the conduction band and valence band. In this process, emission of a photon will occur only if the conduction band has an electron, and the valence band a hole at the same moment of time. In the conditions of thermodynamic equilibrium, the distribution of electrons in the conduction and

the valence band is given by the Fermi-Dirac function

$$f_c = \frac{1}{1+\exp\left(\frac{E_c - E_{Fn}}{kT}\right)}$$
$$f_v = \frac{1}{1+\exp\left(\frac{E_v - E_{Fp}}{kT}\right)} \qquad (4.13)$$

where T is absolute temperature; k is the Boltzmann constant; E_{Fn} and E_{Fp} are quasi-Fermi levels for electrons and holes respectively.

The distribution functions f_c and f_v specify the probability that an electron fills a certain energy level E. The functions $1 - f_c$ and $1 - f_v$ then define the probability that the given level is free or, which is the same, occupied by a hole. So, the probability that an electron and a hole are simultaneously present in the conduction and the valence band, respectively, is equal to the product $f_c\,(1 - f_v)$. The probability of the reverse case, when an electron is in the valence band and a hole in the conduction band, is equal to $f_v\,(1 - f_c)$. Denote the probability of a transition between the conduction and the valence band by B_{em} for photon emission and by B_{abs} for photon absorption. The rates of emission and absorption can then be written as

$$v_{em} = AB_{em} f_c\,(1 - f_v)\,Q \qquad (4.14a)$$
$$v_{abs} = AB_{abs} f_v\,(1 - f_c)\,Q \qquad (4.14b)$$

where A is the density of states in the bands; and Q is the radiated energy.

The condition of light amplification takes the form of inequality $v_{em} > v_{abs}$. From expressions (4.13) and (4.14) it is then easy to derive the condition of population inversion for the case of interband transitions, i.e. the condition of the lasing action:

$$E_{Fn} - E_{Fp} > E_c - E_v \qquad (4.15)$$

Injection lasers attain the population inversion at high levels of injection. Besides, for the condition (4.15) to be fulfilled, the material of at least one of the regions, either the p or the n region, must be degenerate. Fig. 4.4*a* displays the energy diagram for the pn junction of an injection laser in thermodynamic equilibrium, when there is no bias on the pn junction. Fig. 4.4*b* illustrates the diagram for the forward biased pn junction at the current density above threshold.

Along with radiative recombination, other recombination mechanisms come into play, which do not provide emission of radiation. These nonradiative mechanisms include the recombination through structural defects and uncontrollable impurities, Auger (collisional) recombination and others. The nonradiative recom-

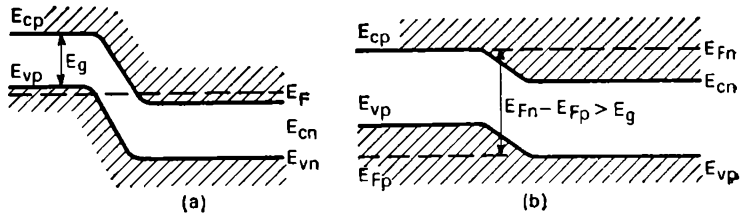


Fig. 4.4. Band diagrams of a *pn* junction (*a*) in equilibrium state; (*b*) in lasing condition

bination of a free carrier evolves a photon of too small an energy. This photon is lost since it does not contribute to the lasing action to produce a useful coherent flux. The laser efficiency obviously rises with the share of acts of radiative recombination as against that of nonradiative acts.

Whatever the mechanism of recombination, the emission wavelength can be defined by the expression

$$1.23/(E_2 - E_1) = 1.23/(E_g - \Delta E) \qquad (4.16a)$$

where E_g is the gap energy; and ΔE is the impurity activation energy. The condition of emission from a direct-gap semiconductor then assumes the form

$$E_P = h\nu \approx E_g \qquad (4.16b)$$

For this, the injected (excited) electrons must take up a sufficient energy from the electric field, enough to satisfy the condition

$$V_{pn} > E_g/q \qquad (4.16c)$$

where V_{pn} is the forward voltage across the *pn* junction.

Thus, the condition (4.16c) along with (4.15) points out that the laser effect can only be produced in degenerate semiconductors.

4.3.2. Injection lasers. Here the specular cuts of the crystal cleaved along its faces normal to the *pn* junction plane function as a resonant cavity. When a sufficiently large forward current flows through the *pn* junction, the structure begins to emit coherent radiation.

The first injection lasers used gallium arsenide as an active medium. A typical GaAs laser has the shape of a rectangular parallelepiped with sides from fractions of 1 mm to 1 mm in length (Fig. 4.5). The laser consists of polished end surfaces *1*, *p* region *2*, *n* region *3*, electric conductors *4*, gold-coated molybdenum plate *5*, and *pn* junction region *6* (shaded stripe). Two end faces act as mirrors of the cavity. The refractory index of GaAs is rather high, and uncoated polished faces reflect about 35% of the incident radiation. Two other faces perpendicular to the *pn* junction are slightly beveled as illustrated in Fig. 4.5.

GaAs lasers can operate both in the pulsed and in the CW mode at low temperatures, and only in the pulsed mode at room tempe-

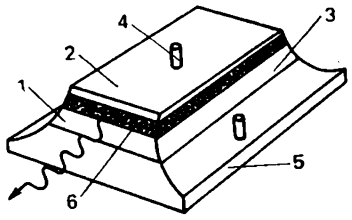


Fig. 4.5. Injection semiconductor laser

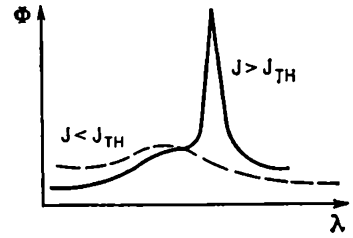


Fig. 4.6. Emission spectra of a laser below and above threshold

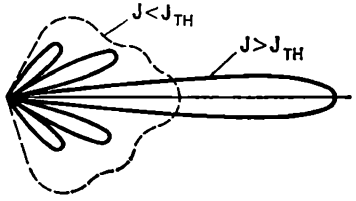


Fig. 4.7. Radiation patterns of a laser below and above threshold

rature. The process of lasing depends on the injection current density J. On applying the forward voltage across the pn junction, the difference between the electron and hole quasi-Fermi levels, $E_{Fn} - E_{Fp}$, begins to rise with J. The photons emitted in the course of carrier recombination differ in energy and in the direction of propagation. But there are photons which propagate in the junction plane. As they collide with excited electrons and bounce back from cavity mirrors, the photons trigger stimulated emission (see Fig. 4.2). The number of such photons increases with the forward voltage and, hence, the current density through the junction. When J climbs to a certain threshold current density J_{TH}, the condition of population inversion begins to prevail. The process of growth of stimulated emission gives way to a qualitatively new process that marks the onset of the laser mode. The emission spectrum now narrows sharply (Fig. 4.6) and the radiation pattern becomes better (Fig. 4.7). The laser emits a coherent flux. In injection lasers, J_{TH} ranges between 10^3 and 10^4 A/cm^2 ($T = 300$ K).

In addition to J_{TH}, the quantum efficiency is the important parameter. The external quantum efficiency is the ratio of photons N_{em} emitted in unit time to the number of electron-hole pairs, N_{pn}, traversing the pn junction for the same time:

$$\eta = \frac{N_{em}}{N_{pn}} \eta_e \eta_o \qquad (4.17)$$

where η_e is the internal quantum efficiency; and η_o is the optical power efficiency defined as the ratio of the radiated power to the power generated in the crystal (see Ch. 1).

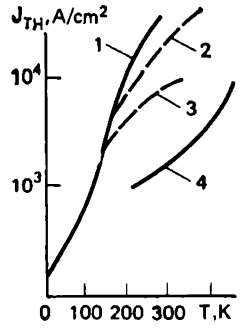


Fig. 4.8. Threshold current density versus temperature

Depending on the amount of impurities they contain homogeneous p and n semiconductors, gallium arsenide, for example, can display four types of emission, or four emission lines. The first type involves the transition of a free hole to the conduction band or to a shallow donor level, the second results when a free electron makes the transition to an acceptor level, the third when the electron jumps from the donor to the acceptor level, and the fourth when the electron decays from the donor level and falls into the valence band. The third type of emission prevails in the p-type semiconductor.

In injection GaAs lasers, it is commonly the p region that emits radiation. The emission line here is analogous to the main (third) line in the homogeneous p-type material.

The laser performance heavily depends on temperature, which causes a shift of its emission spectrum since, first, it changes the refractive index of the laser material with the attendant change in the natural frequency of the cavity and, second, displaces the peak of the luminescence line.

Temperature also strongly affects the threshold current density. Fig. 4.8 illustrates the plots of J_{TH} versus temperature for various injection lasers. For diffused or epitaxial-grown lasers, the values of J_{TH} at room temperature may differ by a factor of a few units and range between 2×10^4 and 5×10^4 A/cm² (curves *1* and *2*). This circumstance inhibits continuous-wave lasing at room temperature. Conversion to heterostructures made it possible to reduce J_{TH} and appreciably improve the performance of injection lasers (curves *3* and *4*).

4.3.3. Heterojunction lasers. The energy diagrams of heterostructures (see Ch. 1) exhibit different potential barriers for counterflows of electrons and holes, thus promoting a unidirectional injection of carriers from the wide-gap emitter to the narrow-gap base. The concentration of carriers injected into the base can be a few orders of magnitude higher than the equilibrium concentration of carriers in the emitter region. The emitter and the base layer of the heterostructure differ in optical properties, particularly in refractive index n, because the gap energy of the emitter is much higher than that of the base. There is no need to dope the semiconductor of a heterolaser until its degeneration since the energy gap difference encourages population inversion of energy levels. A decreased level of doping is conducive to a lower loss by nonradiative recombination and higher internal quantum efficiency.

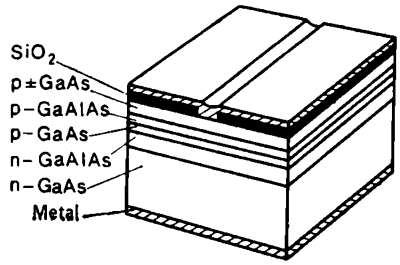


Fig. 4.9. Injection-type double-heterojunction laser

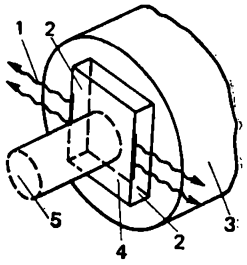


Fig. 4.10. Electron-beam-pumped semiconductor laser

Unidirectional injection specific to heterojunctions tends to accumulate excess carriers in the active medium region and to almost fully inhibit their penetration into the emitter. The waveguide effect also plays a positive role since it favors the concentration of the wave inside the optically denser medium layer of the structure. In the final analysis, heterolasers have tenths of the value of J_{TH} required for homojunction lasers and higher efficiency, which, in turn, enables them to lase continuously at room temperature.

Multilayer-heterojunction lasers offer the highest promise. Fig. 4.9 demonstrates a double-heterojunction (DH) stripe-geometry laser. For the best designs of DH GaAs-GaAlAs lasers, the threshold current density J_{TH} at room temperature is now lower than 10^3 A/cm^2, which is a factor of about 100 smaller than that for homojunction lasers.

4.3.4. Electron-beam lasers. These lasers use a high-energy electron flux of tens and hundreds of keV to excite a semiconductor. As they penetrate into the crystal bulk, the pumped electrons excite a fraction of valence electrons to higher energy levels of the conduction band. These excited electrons in turn give off energy to other lattice atoms, thus inducing a photon avalanche which decays with distance away from the source into the semiconductor medium.

An electron-beam-pumped laser resembles a cathode-ray tube in design, which operates at tens of kilovolts to excite a semiconductor placed in a tube. Fig. 4.10 illustrates the principle of electron-beam pumping. Here, *1* designates coherent emission; *2*, reflecting surfaces; *3*, substrate; *4*, semiconductor; and *5*, electron flow. The electron flow of 20 keV and over impinges on the flat face of the semiconductor and generates in its thin surface layer a large number of electron-hole pairs, about 10 pairs per electron. The carriers so produced collect at the bottom of the conduction band (electrons) and at the top of the valence band (holes) and recombine there. The coherent flux escapes the semiconductor chip in the plane normal to the electron beam direction. The reflecting semiconductor faces serve as mirrors of the open resonant cavity. The thickness of the active layer depends on the energy of electrons and can reach tenths of a millimeter.

The type of laser pumping illustrated in Fig. 4.10, where the electrons travel in the direction normal to the axis of the laser cavity is known as transverse pumping. In longitudinal pumping, the electron flow moves along the cavity mirrors. Electron-beam pumping enables the construction of laser systems with a power of up to 1 MW in a pulse.

Electron-beam lasers are not free from disadvantages which limit the area of application of these devices. The principal shortcomings are the need for a highly evacuated space, large size, low throughput optical efficiency (not above 1%), complex and cumbersome supply system (its volume and mass are tens of times as large as those of the laser proper). Nevertheless, this lasing emitter is most suitable for projection large-screen color television (with screen areas up to 10 m^2), superhigh-speed readers of computer-processed information, high-speed holographic storages, and generators of light pulses 10^{-12} s in duration.

4.4. Laser Types

4.4.1. Gas lasers. These devices find use in optoelectronics primarily because they offer a high degree of optical coherence; the normalized change of the emission wavelength, $\Delta\lambda/\lambda_{max}$, lies in the range 10^{-6}-10^{-9} as against 10^{-4} for other types. A low value of $\Delta\lambda$ is accounted for by a weak interaction of atoms and molecules in a highly rarefied gaseous medium. The laser gain K_l in rarefied gases is low. A gas laser must then be large in size to make quantum amplification efficient. But this disadvantage proves to be an asset since large linear dimensions afford a low beam divergence θ; as is obvious from expression (4.19), a resonant cavity not less than tens of cm in length can provide a beam whose angle of divergence is $\theta < 0.01°$ as against $\theta > 1°$ for other types of laser.

Thus, gas lasers are more preferable where high coherence and low divergence are important factors. For example, gas lasers are now indispensable for laser communication systems used to transfer a large body of information.

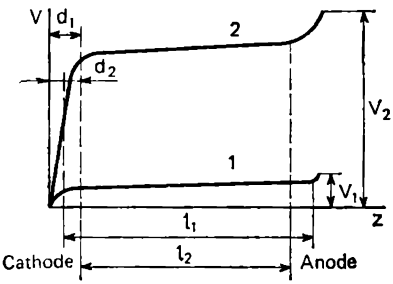


Fig. 4.11. Potential distribution in arc (*1*) and glow (*2*) discharges

Consider an example of the gas-discharge laser to gain insight into the principle of its operation. This laser uses a rarefied medium at a pressure of 100 to 1 000 MPa. The active medium here is plasma produced by a discharge of the arc or glow type initiated in a gas or gas mixture. Fig. 4.11 illustrates the

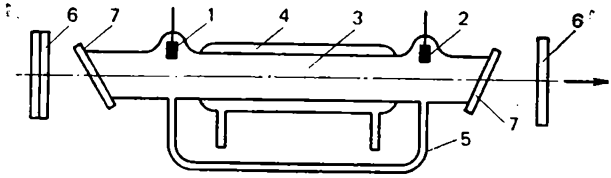


Fig. 4.12. Schematic form of an arc-excited gas laser

potential distribution along the discharge gap length for arc and glow discharges. Here index *1* (and curve *1*) relate to the arc discharge, and index *2* (and curve *2*) to the glow discharge; V is the operating voltage, i.e. the potential difference between anode and cathode; d is the length of a near-cathode discharge region; and l is the length of the discharge region called a column where discharge characteristics are most stable and the field strength is constant over its entire length. This is the operating region for a gas-discharge laser. From comparison of the key parameters of both discharge types, we find that $d_1 \ll d_2$, $V_1 \ll V_2$, and $l_1 > l_2$. Some parameters of arc and glow discharges are given in Table 4.1.

Table 4.1. Parameters of Arc and Glow Discharges

Discharge	Operating voltage, V	Current density, A/cm²	Temperature, K	Electron concentration, cm⁻³	Relative ionization, %
Arc	$10\text{-}10^2$	$10^2\text{-}10^3$	10^3	10^{13}	1
Glow	10^3	$10^{-3}\text{-}10^{-1}$	3×10^2	$10^9\text{-}10^{11}$	10^{-2}

It is seen from the table that the arc discharge is a high-current high-temperature discharge with a high degree of ionization; the glow discharge is a weak-current low-temperature low-ionization discharge.

Figure 4.12 illustrates the schematic diagram of an arc-excited argon (Ar) laser. The setup includes an anode *1*, cathode *2*, active medium *3*, water cooling arrangement *4*, bypass channel *5*, fully reflective and output mirrors *6*, and output windows *7* of the gas-discharge tube. The bypass channel serves to level off the pressure along the tube length: in the absence of this channel the gas accumulates in the anode section of the tube soon after initiating the arc discharge. This type of laser is rather costly and bulky.

The helium-neon (He-Ne) laser excited by a glow discharge is most popular. The discharge sustained in the tube excites helium atoms which collide with neon atoms raised exactly to the same levels as the helium atoms. Neon atoms take off energy from helium

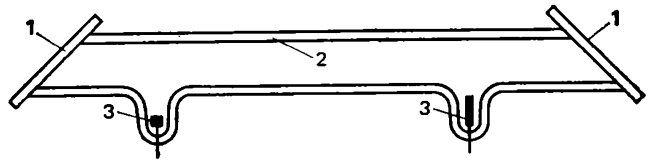


Fig. 4.13. Glow-discharge-excited gas laser tube

atoms and provide coherent emission at a wavelength of 0.633 μm.

The main unit of a He-Ne laser is a glass or a quartz tube *2* a few millimeters in diameter and a few centimeters up to a few meters in length, which is pumped to a high vacuum and then filled with a He-Ne mixture (Fig. 4.13). The supply voltage is fed to electrodes *3* to build up and sustain the discharge in the gas mixture. The tube ends are covered with flat glass or quartz plates *1*. Reflection from the flat plate depends on both the angle of incidence and the polarization of radiation. If the plane of polarization coincides with the plane of incidence of radiation on the plate, then at a definite angle called the Brewster angle, the radiated wave will pass through the plate without any reflection.

Gas lasers are in the main impracticable in the field of microoptics for their large overall dimensions, gas-filled tubes, high-voltage supply, low efficiency, and complex pumping devices. But, as noted above, where high coherence and low beam divergence are important, while the mass and size requirements are not decisive, gas lasers are the best choice.

4.4.2. Solid-state lasers. The active medium used in these lasers is a crystalline or amorphous dielectric material having luminescence centers. The main type of solid-state laser employed for small-size optoelectronic devices compatible with microcircuits is an yttrium aluminum garnet (YAG) laser. The garnet $Y_3Al_5O_{12}$ has neodymium ions Nd^{+++} which substitute for a fraction of yttrium atoms in its crystal lattice. The crystal is mechanically strong, highly heat-conducting, and homogeneous as regards its optical and physical parameters. The laser uses optical pumping to induce coherent emission.

Structurally, a solid laser contains a laser rod with polished mirror faces to perform the function of a resonant cavity, a flash or a continuous pump lamp to illuminate the rod, and light-focusing system to uniformly illuminate the rod surface with a minimum of its heating.

Since the concentration of active centers in the crystals of solid lasers is high, the energy of coherent emission proves to be tens of times as intensive as that of other lasers. Optical pumping gives a higher efficiency than the gas discharge. Solid lasers feature improved temperature and radiation stability and also higher mechanical strength.

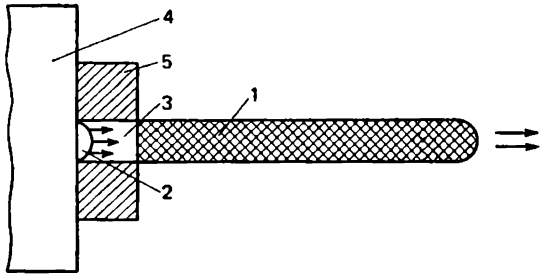


Fig. 4.14. Solid-state laser structure

Yet, optical pumping presents a drawback since it involves an additional stage of energy conversion (light-electricity-laser emission) and does not allow for a higher output. Modern elements of the optical pump have a short life and decrease the reliability of the laser as a whole. The principle of excitation of solid lasers does not permit internal modulation of laser emission.

Figure 4.14 illustrates the schematic diagram of a solid laser which includes a laser rod *1*, IR diode *2* as a pump, optical medium *3*, heat sink *4*, and laser head holder *5*. Replacement of an incandescent lamp by an IR diode makes it possible to reduce laser dimensions and increase the efficiency. Solid lasers usually employ GaAlAs IR diodes emitting at λ_{max} of about 0.81 μm, which coincides with the resonance absorption band of a laser, thus affording an increase in the output efficiency; incandescent lamps have a smeared emission spectrum so that only a portion of this spectrum is in resonance with the laser spectrum.

4.4.3. Comparison of lasers. Table 4.2 lists the basic parameters of lasers used today. The data enable us to conclude that only the

Table 4.2. **Parameters of Small-Size Lasers**

Type	Size, cm	Efficiency, %	Φ_e, mW	λ_{max}, μm	$\Delta\lambda/\lambda_{max}$	θ, °	V_S, V
Gas	10	10^{-1}-10^{-2}	0.1-10	0.63	10^{-6}-10^{-9}	0.03-0.15	10^3-10^4
Solid	1	1-2	10-10^4	1.06	10^{-4}	1	up to 10^2
Semiconductor	0.1	10-20	10-10^2	0.8-0.9	2×10^{-3}	10	1.5-3

injection semiconductor laser is a sufficiently universal coherent source. The advantages of semiconductor lasers are the following: (1) small size and mass and high optical amplification with K_l ranging from 10^3 to 10^4 cm^{-1}; (2) high output efficiency; (3) simple pumping that does not require high supply voltages and powers; (4) high speed of response; (5) emission at the specified wavelength in a wide

range, using for the purpose semiconductors with the requisite value of gap energy; and (6) manufacturing and operational compatibility with elements of integrated optics.

The disadvantages inherent in modern semiconductor lasers are: (1) low values of both the emission wavelength change and the beam divergence θ due to an enhanced density of the active material, short length of the resonator, and small exit aperture; and (2) short service life, which does not exceed 10^2-10^3 hours, though theory confirms that the life of injection lasers can be longer than 10^5 hours.

A decrease in the life of actual devices results from gradual degradation (aging) of laser materials under the action of high current densities and intensive radiant and thermal fluxes typical of semiconductor lasers. The main cause responsible for degradation is an increase in the concentration of nonradiative centers in the active medium due to interstitial intrusion of uncontrollable impurities and formation of other defects. Besides, degradation manifests itself as a decrease in the activity of radiative centers and increase in surface recombination.

The basic steps taken to reduce the rate of degradation and to prolong the life of semiconductor lasers amount to (1) selecting materials uniform in structure and in electrical, thermal, and physical parameters; (2) perfecting crystal geometry; and (3) improving crystal mounting and heat sink.

Solid-state lasers offer the highest promise in the areas requiring high pulse powers in combination with low divergence, for example, in laser location, object lighting, and long-range optical communication. Gas lasers remain beyond competition in all devices and systems where coherence is a decisive factor.

4.5. Laser Modulators

4.5.1. Laser modulation. Modulation of laser emission is the process of modifying a parameter or parameters of the beam in the appropriate manner in time and/or space. The pattern of changes in the modified parameter conforms to the pattern of changes in the signal (intelligence) transmitted by the beam. In other words, the laser beam serves as a carrier of information impressed on it by appropriately changing its parameter. For example, if the intelligence is embedded in the radiant intensity of the carrier (beam), i.e. if the useful signal modifies the carrier intensity, the intensity modulation results. Frequency modulation implies a change in the carrier frequency by the intelligence signal; phase modulation involves changes in the carrier phase; polarization modulation is the process of changing the direction of polarization. The devices used to modulate a useful signal on to the carrier fall into following basic groups.

1. Modulators, which change one or a few parameters of laser emission in time, such as intensity, phase, frequency, and polarization.
2. Deflectors, which appropriately change the beam position in time.
3. Space-time modulators (STMs), which change in time the spatial distribution of beam intensity, phase or polarization. STMs combine the functions of a modulator with the functions of a deflector, thereby 'loading' more information on the carrier.

4.5.2. Physical principles of laser modulation. Optical modulation techniques largely rely on electrooptic and magnetooptic (Faraday) effects, and also various photoeffects.

Electrooptic effects induce optical anisotropy in a substance subjected to an external electric field and, hence, change the refractive index of the material. Optical anisotropy appears from changes in the relative permittivity of the material exposed to the electric field.

Electrooptic effects usually entail the phenomenon of birefrigence—splitting of the propagating beam into two ordinary and extraordinary rays. These two rays travel with different speeds and have different directions of polarization. In wave-transmission crystals the refractive indexes in two mutually perpendicular directions x, y of light propagation are generally different. Denote the refractive index along each axis by n_x and n_y respectively. Crystals in which the refractive indexes along these two axes differ in value, $n_x \neq n_y$, are known as biaxial crystals. Crystals in which the directions x and y are optically homogeneous, i.e. $n_x \approx n_y = n_0$, are called uniaxial. In uniaxial crystals the refractive index for the ordinary wave is equal to $n_0 = n_x = n_y$; for the extraordinary wave, the refractive index is $n_e = n_z$.

The speed of light propagating along the z axis in a uniaxial crystal is independent of the character of polarization. An externally applied field disturbs the equality of n_x and n_y and the crystal becomes biaxial. The speed of propagation of light waves polarized along the x and y axes also begins to change.

The index of refraction for an ordinary wave traveling along the z axis varies linearly with the electric field strength E:

$$n_0 (E) = n_0 + r_P E \tag{4.18}$$

where r_P is the electrooptic Pockels constant; and n_0 is the refractive index of a material not exposed to the field.

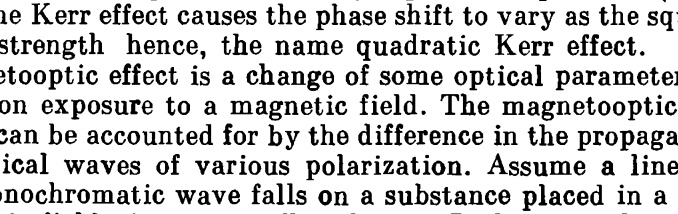
This change in the refractive index by an amount proportionate to the field strength is the Pockels effect, or linear electrooptic effect. The externally applied field imparts the uniaxial crystal the properties of the biaxial crystal, so that it becomes optically anisotropic due to a change in its refractive index. As the light wave in such a crystal propagates through a path of length l, there appears a phase difference between the ordinary and the extraordinary

ray:

$$\Delta\varphi_{0e} = 2\pi n_0^3 r_P E l/\lambda \qquad (4.19)$$

As the wave goes on moving deeper into the crystal bulk, the phase difference between the optical signals begins to change. The output and input signals thus differ in polarization. Depending on the path length covered by the wave in the crystal and, hence, the phase difference $\Delta\varphi_{0e}$, the polarization of the output signal will vary as illustrated in Table 4.3.

Table 4.3. Polarization of the Output Beam as a Function of the Phase Shift Between the Ordinary and Extraordinary Rays

$\Delta\varphi_{0e}$	0	$\pi/4$	$\pi/2$	$3\pi/4$	π	$5\pi/4$	$3\pi/2$	$7\pi/4$	2π
Polarization									

As regards the mutual orientation of the beam direction z and the field strength direction, the Pockels effect can be either longitudinal ($z \parallel E$) or transverse ($z \perp E$).

Optoelectronics also uses an electrooptic Kerr effect. Here the refractive index varies with the field strength according to the relation

$$n_0(E) = n_0 + r_K E^2 \qquad (4.20)$$

where r_K is the Kerr constant, which accounts for the material properties, temperature, and signal wavelength.

The phase shift appearing between optical signals after field application is written as

$$\Delta\varphi = 2\pi r_K l E^2 \qquad (4.21)$$

where l is the path length traversed by signals. Expression (4.20) shows that the Kerr effect causes the phase shift to vary as the square of the field strength hence, the name quadratic Kerr effect.

The magnetooptic effect is a change of some optical parameters of a substance on exposure to a magnetic field. The magnetooptic Faraday effect can be accounted for by the difference in the propagation speed of optical waves of various polarization. Assume a linearly polarized monochromatic wave falls on a substance placed in a constant magnetic field of magnetic flux density B; the wave direction coincides with the field direction. As is known, the linearly polarized wave can be represented as a sum of two waves differing in polarization. The magnetic field makes the refractive indexes n_1 and n_2 for these two waves different from each other. So, as the light traverses a distance l in the crystal, the two waves will differ in phase by

$$\Delta\varphi = (\omega l/c)\ (n_1 - n_2) \qquad (4.22)$$

where $n_1 - n_2$ is the difference in refractive indexes that is proportional to B.

Among photoeffects responsible for changes in optical parameters of materials, we can single out photochromatic, photocrystalline, and photoconductive effects.

The photochromatic effect shows up as a change in the coloration or transparency of inorganic glasses having special additions and organic polymers by the action of the light flux in a short-wavelength region such as UV and short-wavelength visible radiation regions. To bring the material back to its original state, we have to heat it or expose to infrared rays.

The photocrystalline effect appears in amorphous semiconductors: a highly intensive beam causes a semiconductor crystallize and thus change its index of refraction.

The photoconductive effect is given detailed treatment in Ch. 2. A change in the conductivity under illumination changes the optical properties of a material, particularly its refractive index. This feature is taken advantage of for laser modulation.

4.5.3. Optical modulators. As mentioned earlier, modulation of a beam (carrier) impresses the intelligence (useful signal) on the carrier. According to equation (4.1) for a light wave, the intelligence can be embedded in the amplitude, phase, and polarization vector of the carrier to be routed to a photodetector. Modern photodetectors only respond to the intensity of radiation, for which reason intensity modulators are the most extensively used devices. An optical signal derived from any other type of modulator requires modulation in intensity before applying it to a PD.

A modulator provides external modulation of a laser beam by changing its parameters. Internal modulation inside the laser is also possible. This is done by changing the parameters of the active medium or cavity. In semiconductor lasers, internal modulation arises by way of changing the pump operation in the desired manner. In gas lasers, internal modulation can be put into effect by switching the cavity Q. It should be pointed out, however, that combining in one device the functions of lasing and modulation generally impairs the parameters of laser radiation because such a device develops side effects which degrade beam coherence. In semiconductor lasers which most readily lend themselves to internal intensity control, internal modulation commonly causes a growth of the lasing threshold and impairs the set of excited modes.

Let us look at the structure of an electrooptic modulator (Fig. 4.15). The building blocks of the modulator are two crystals of the same size. Crystallographic axes of the crystals are normal to each other to compensate for the effect of temperature on the modulation process: temperature changes subtract, as it were, when the optical signal sequentially passes through both crystals. As seen from the figure,

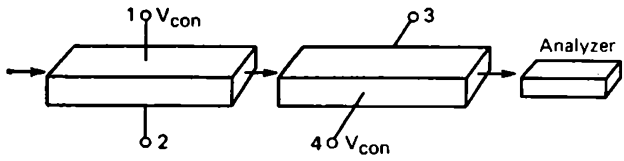


Fig. 4.15. Block diagram of an electrooptic modulator

the modulator uses the transverse Pockels effect, $z \perp E$. Changing the control voltage V_{con}, i.e. electric field strength E, adjusts the phase of the output signal with respect to the input signal phase. The output analyzer converts the phase changes into the changes of radiation intensity.

Disregarding the loss by absorption in the crystal, the radiation intensity E_O at the modulator output takes the form

$$E_O = E_I \sin^2 \left(\frac{\pi}{2} \frac{V_{con}}{V_{\lambda/2}} \right) \tag{4.23}$$

where V_{con} is the control voltage; and $V_{\lambda/2}$ is the half-wave control voltage.

The half-wave voltage is an important parameter of the modulator. This is a control voltage at which the modulator transmittance can vary from minimum to maximum, and vice versa. In this case, two rays shift one relative to the other by half the wavelength, i.e. the phase shift is $\Delta\varphi = \pi$. For the above modulator the half-wave voltage is

$$V_{\lambda/2} = (\lambda/2n_0^3 r_P)\,(d/l) \tag{4.24}$$

The values of $V_{\lambda/2}$ for various modulators range from hundreds of volts to units of kilovolts. Such high values of control voltage are a serious limitation to the usefulness of electrooptic modulators.

The plots of output radiation intensity E_O versus control voltage V_{con} in relative units are shown in Fig. 4.16. This is a control characteristic of the modulator. The modulation process illustrated in Fig. 4.16*a* occurs without the dc bias voltage V_b, which is otherwise fed to the same electrodes as those used to supply V_{con}.

For the case in hand, the modulator effectiveness is low: the output intensity practically remains the same with appreciable variations in the control voltage. This effectiveness can be defined in terms of the degree of modulation (modulation factor)

$$m = E_{max}/E_{min} \tag{4.25}$$

When the bias voltage V_b is zero, we have $m \to 0$. A dc bias $V_b = V_{\lambda/2}$ applied to the modulator (Fig. 4.16*b*) appreciably augments both the amplitude of the ac component and the degree of modula-

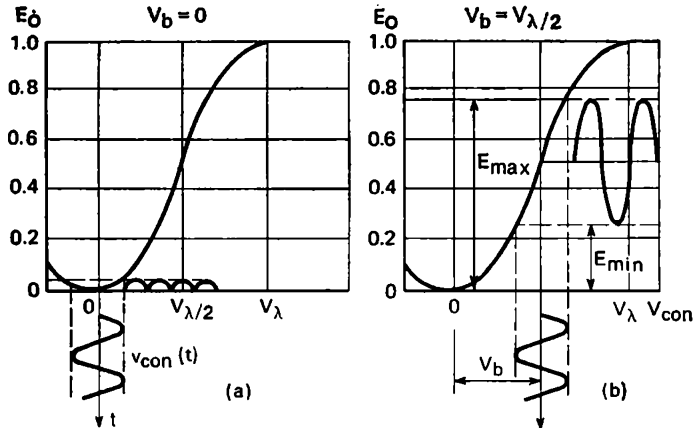


Fig. 4.16. Control characteristics of a radiant-intensity modulator without bias (*a*) and with bias (*b*)

tion. The modulator operates in the steep region of its control characteristic.

The upper frequency limit of such a device ranges into 10^8-10^9 Hz. A figure of merit sometimes used to assess the modulator performance is the quality factor Q_M defined as the ratio of the limiting frequency to the control power:

$$Q_M = f_{\text{lim}}/P_{con} \tag{4.26}$$

The Q value for electrooptic modulators is close to 10 MHz/mW.

Magnetooptic materials such as garnet ferrites or chromium tribromide can also be used to perform the function of an active medium. A magnetooptic modulator depends for its operation on the Faraday effect (Fig. 4.17). As the light propagates through the active medium placed in a magnetic field, the plane of polarization of the beam rotates through a certain angle. The analyzer then converts the changes in the direction of polarization into changes in the amplitude.

Magnetooptic modulators are much inferior to electrooptic modulators in speed (the frequency limit does not commonly exceed 10^4 Hz). Besides they require a high magnetic field strength for their control. Other limitations to the usefulness of these devices are a low degree of modulation and heavy absorption of optical radiation.

4.5.4. Thin-film and semiconductor modulators. Thin-film modulators relying on the same physical effects as those applied in cavity modulators are the most promising devices for use in microoptic systems. Electrooptic thin-film modulators utilize niobate and tantalate of lithium and also solid solutions based on these materials to

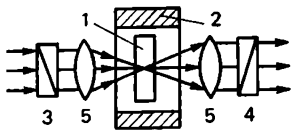


Fig. 4.17. Block diagram of a magnetooptic modulator (*1*) active medium; (*2*) coil; (3) polarizer; (*4*) analyzer; (*5*) lenses

perform the function of an active medium. Magnetooptic thin-film modulators afford an increase in the limiting frequency up to 10^8 Hz as against 10^4 Hz for cavity-type counterparts. Both types of thin-film modulators require low control voltages (units of volts), which makes them compatible with microelectronic devices.

Semiconductor modulators look exceptionally promising for microoptics. A linear electrooptic effect appears in a semiconductor *pn* junction as the externally applied reverse voltage changes the free carrier concentration in the space charge region. The result is the modulation of the relative permittivity and, hence, a change in the refractive index *n*. This effect is difficult to apply for cavity modulators because it would be necessary to guide the beam to a narrow region along the *pn* junction. Thin-film structures readily comply with this requirement. In modulators using GaAs and GaP semiconductors, the control voltage is as low as units of volts.

Semiconductors absorb radiation as free electrons make both intraband and interband transitions from the valence to the conduction band. These two types of absorption are put to practical uses for modulation of the light intensity. The radiation incident on a semiconductor plate decays exponentially with a distance equal to the penetration depth χ_0 for the given semiconductor. If χ_0 varies from χ_{min} to χ_{max}, the range of changes in the light intensity becomes

$$E_{min} = E_0 \exp\left(-\frac{l}{\chi_{min}}\right) \leqslant E \leqslant E_0 \exp\left(-\frac{l}{\chi_{max}}\right) = E_{max} \qquad (4.27)$$

where l is the semiconductor thickness. The degree of modulation is then equal to

$$m = \frac{E_{max}}{E_{min}} = \frac{\exp(-l/\chi_{min})}{\exp(-l/\chi_{max})} \qquad (4.28)$$

The degree of modulation tends to unity as the thickness l goes into infinity. But an increase in l sharply lowers the output intensity, i.e. decreases the modulator efficiency.

Figure 4.18 illustrates the modulation factor as a function of the χ_{max} to χ_{min} ratio. As seen, m reaches an acceptable value of 0.6 to 0.7 as the absorptance changes by a factor of about 10.

Semiconductor modulators operating on the principle of control of light absorption by free carriers adjust the concentration of carriers by injecting them through the *pn* junction. A wider range of changes in the free carrier concentration extends the limits of changes in the optical density of the semiconductor and, consequently,

adds to the degree of modulation. However, a change in the carrier concentration affects other electrophysical parameters of the structure, for example, prolongs the lifetime of carriers so that the upper frequency limit drops to 10^5-10^6 Hz. The advantages of these modulators are that they operate at low control voltages, are simple in design, reliable in operation, and adaptable to srteamlined manufacture.

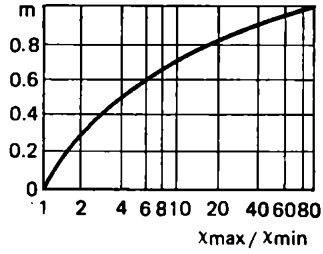


Fig. 4.18. Degree of modulation versus absorption

4.5.5. Deflectors. The deflector is a device intended to change the laser beam position in space. Electrooptic and acoustooptic deflectors have found wide uses in holographic recording and readout systems.

A typical block diagram of the electrooptic deflector appears in Fig. 4.19. The deflector is a multistage device. Each stage consists of an optical polarization modulator *1* and a birefrigent crystal *2*. The modulator smoothly changes the laser polarization in the desired manner. The crystal in each stage splits the beam into two rays.

Depending on the set of control voltages, the deflector can yield 2^i discrete positions of the output beam in space, where i is the number of stages. The total number of spatial positions of a beam that is achievable in the electrooptic deflector is given by

$$N = \varphi D/\lambda \tag{4.29}$$

where φ is a maximum angle of deflection of the beam; D is the deflector aperture; and λ is the wavelength.

What distinguishes this deflector is its high speed; in a device with N equal to 1 000, the switching time is about 0.1 μs. With an increase in N and thus in the number of stages, the speed diminishes. An optimum design is the one that requires a minimum of control power P_{con} at the shortest switching time t_{sw}. An optimization approach is to focus the laser beam traveling through the deflector and to reduce the parasitic capacitance of the crystal. As with the modulator, the figure of merit that characterizes the performance of the deflector is its Q factor

$$Q_D = \frac{N}{t_{sw} P_{con}} \approx \frac{n^7 r_P^n}{\varepsilon \lambda^3} \tag{4.30}$$

where n is the refractive index; ε is the relative permittivity; and r_P is the electrooptic Pockels constant.

Deflectors use the same materials as modulators, for example, lithium niobate $LiNbO_3$ and others. A large number of crystals in the deflector brings forth more stringent requirements on both the

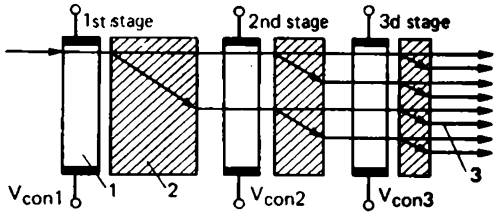


Fig. 4.19. Block diagram of an electrooptic deflector

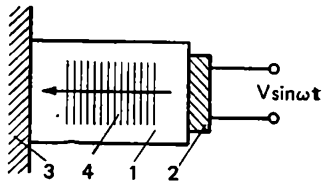


Fig. 4.20. Acoustooptic deflector

starting materials and the design on the whole; high transparency and also exceptionally accurate polishing and mounting of each crystal are the basic requirements.

Acoustooptic deflectors also find application. Fig. 4.20 illustrates the block diagram of this type of deflector which includes an active medium *1*, piezoelectric transducer *2* (exciter), acoustic absorber *3*, and excited diffraction grating *4*. The deflector operates on the following principle. The acoustic wave produced by the transducer propagates through the active medium and induces compressed and rarefied regions in the medium, thus changing its refractive index. The wave actually forms a certain grating consisting of the regions with different refractive indexes. The period of changes in the refractive index of the grating is equal to the length of the acoustic wave, and the amplitude of changes is proportional to the wave amplitude and dependent on the properties of the active medium. The optical wave incident on this grating undergoes diffraction, i.e. changes its position in space. The spatial position of the wave is controlled by changing the frequency of acoustic vibrations and appropriately adjusting the diffraction grating.

A figure of merit for an acoustooptic material is the ratio of the scattered radiation to the power of acoustic vibrations. As regards this parameter, the most suitable crystalline materials are lead molybdate $PbMoO_4$, paratellurite TeO_2, lithium niobate $LiNbO_3$, and some others.

Obviously, an acoustooptic deflector tunes in to a new frequency (i.e. the diffraction grating acquires a new state) only when the incoming acoustic wave traverses the entire active medium. This circumstance explains why this type of deflector has a low frequency limit, up to 10^6 Hz.

4.5.6. Space-time modulators (STMs). These are either electrically or optically controlled devices. An electrically controlled STM uses electric signals to modify the space-time distribution of any parameter of the laser beam in the required manner. An optically controlled STM changes laser beam parameters by an optical beam applied to its input.

The speed of an STM depends on the relaxation time t_{rel}, which is the time of decay of the transient involving changes in an optical parameter after applying the control signal. When the control signal ceases, the active material begins to recover its optical properties with a certain rate that determines the recovery time t_{rec}. Three types of recovery process are possible: (1) the process in the modulation conditions at which $t_{rec} \approx t_{rel}$; (2) the process in the switching conditions at which $t_{rec} \gg t_{rel}$; (3) the process in the relaxation conditions at which the material does not regain its optical properties, $t_{rec} \to \infty$.

We can single out two methods of electrical control of STMs. The first applies to matrix-type STMs. This type of STM consists of cells, each exposed to an electric field of the specified strength. Here the strength of the control field depends on spatial coordinates. The second method uses an electron beam scanned across the surface of the modulator. The intensity of the beam varies in time according to the control program.

Electrically controlled STMs commonly serve to convert an electric signal to a two-dimensional optical signal. STMs can read out information and subsequently enter it into the optical memory of a computer.

Optically controlled STMs change the parameters of optical signals by optical control signals. There are two main methods of optical control, a matrix-type and a projection-type. In the first, the control beam in the form of pulses varying in amplitude impinges on individual elements of the matrix and changes the optical properties of each element in the prescribed manner. In the second, the control beam carries the image projected on the entire surface of the modulator.

Optically controlled STMs can also record information and read it from optical memories, such as holographic ones noted for reliable storage of data, large capacity, and high speed.

4.6. Fiber-Optic Communication Lines

Modern electrical communications such as telegraphy, telephony, and radio communication are at a high level of development. As mentioned in the introduction to this book, further improvements in the field of these telecommunications involve huge expenses, much effort and time, though the results can hardly live up to expectations. There is an urgent need for qualitative changes in the status of modern communication.

The history of radio and other types of communication has shown that with a decrease in the wavelength of an intelligence signal, the level of atmospheric and industrial noise drops off so that the quality of communication improves. Besides, the use of shorter wavelengths

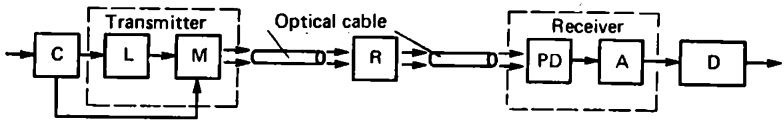


Fig. 4.21. Block diagram of a fiber-optic communication line

permits an increase in the number of possible communication channels.

Lasers opened up possibilities for extending communication bands into a shorter-wavelength region of electromagnetic radiation and stimulated an active search for ways of evolving optical communication lines. But in the course of time it became clear that open laser lines are not reliable enough. Even at a distance of several kilometers, the pouring rain, snow fall or fog can disturb contact. The potential offered by optical communication for long remained out of practical use because there was no optical analog of the electric cable. It was not until the mid 1960s that the first fiber guides appeared, which formed the basis for modern fiber-optic communication lines (FOCLs).

4.6.1. FOCL block diagram. This block diagram illustrated in Fig. 4.21 includes the following basic units: an input coder (C), laser transmitter, fiber optic cable, repeater (R), receiver, and output decoder (D).

The coder converts the initial intelligence to the form convenient for transmission. The basic block of the transmitter is a common semiconductor laser (L). The intelligence encoded in the electric signal enters a modulator (M) which controls the intensity of the laser beam. The optical channel then conveys the modulated optical signal to a photodetector (PD) to convert it to an electric signal, which then arrives at a high-gain amplifier A. The repeater R serves to restore the level of a useful signal transmitted for a long distance. The decoder is a translator of coded signals.

The block diagram of a practical FOCL may differ from that of Fig. 4.21 depending on the purpose the line must serve, its extent, speed, and quality of its units. The repeater can sometimes be absent. In the simplest case, a very short line may consist of a lasing emitter, cable, and photodetector.

The distinguishing features of a FOCL are the following.

1. High noise immunity to external electromagnetic influences and babbles. This advantage is of particular importance for dense communication lines and lines laid near noise sources.

2. Small size and mass due to a low density of starting materials and absence of heavy sheaths. FOCLs have one third to one fifth the mass and size of wire lines. This feature is essential for vehicle-borne equipment.

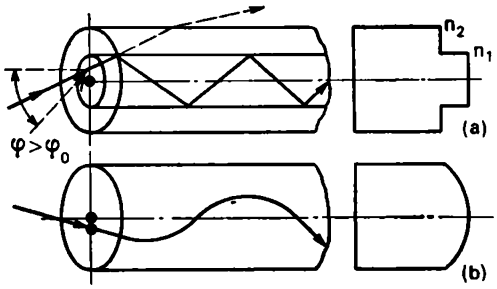


Fig. 4.22. Beam propagation and refractive index changes in light conductors (*a*) two-layer fiber; (*b*) selfoc

3. Wide frequency range, up to 1 GHz and over. As a result, one optical transmission line can simultaneously convey 10^{10} calls or 10^6 TV programs.

4. Privacy of information exchange: the line hardly emits into the environment and any attempt to plug in on the line will as a rule end in damage to the cable.

5. Inherently low cost since here glass, quartz, polymers, and other materials abundant in nature substitute for such costly nonferrous metals as copper and lead. Besides, FOCLs afford ease of manufacture, laying, and maintenance (light guides).

4.6.2. Fiber optic guide. The preceding chapters described the building blocks of FOCLs, such as radiators, photodetectors, and optical modulators. Here focus of attention will be on optical cables composed of individual optical conductors.

Optical energy propagates along a fiber guide by virtue of the effect of total internal reflection (TIR). Consider how this effect manifests itself in a cylindrical two-layer guide (Fig. 4.22*a*). Let the material of the core and external layer (cladding) have refractory indexes n_1 and n_2, respectively. Assume also that $n_1 > n_2$, i.e. the core material is optically denser than the coating. The light beam entering the fiber guide at a small angle to the cylinder axis undergoes TIR: the beam incident on the boundary of the cladding totally reflects backward into the core, and does so over and over again as it travels along the guide without escaping outside. The maximum (critical) angle of deflection of the beam from the axes, at which the TIR is still possible, is given as

$$A_0 = \sin \varphi_0 = \sqrt{n_1^2 - n_2^2} \qquad (4.31)$$

This quantity is called the numerical aperture of a fiber guide, which is taken into account when matching the guide to the photoemitter. The ray impinging on the fiber face at an angle $\varphi > \varphi_0$ (off-aperture ray) undergoes not only reflection but also refraction as it

interacts with the cladding; a fraction of radiation multiply encounters the core-cladding interface and is finally scattered away from the guide.

The beam can also propagate along a fiber guide where the refractive index decreases away from the center gradually. The beam falling on the face of such a guide is deflected and focused near the axial line (Fig. 4.22*b*). Any section of the guide acts as a short focus lens, thus producing the effect of self-focusing. This type of guide is known as a selfoc, or gradient conductor (graded-index fiber).

The loss of radiation traveling along the guide should be at a minimum. This loss, called transmission loss, is defined in terms of optical signal attenuation B:

$$B = \frac{1}{l} 10 \log \frac{P_I}{P_O} \tag{4.32}$$

where P_I is the power of input optical signal; P_O is the power of an optical signal at the exit of the guide; and l is the guide length.

The unit of measure used to estimate the attenuation of an optical signal is the decibel per kilometer, dB/km. Modern light conductors have a high fiber purity, so that the signal attenuation is equal to 0.5 dB/km at a wavelength of 1.2 μm, which is close to the theoretical limit for quartz fiber guides. For comparison, let us note that such a conductor 1 km long is more transparent than thoroughly washed window glass.

The transmission loss in a fiber guide is due to absorption and scattering of optical radiation. The absorption loss is caused by impurity ions, largely by metal ions of the copper-chromium group (fourth group). The intensity of absorption due to metal ions depends on the spectrum of the transmitted signal and many other factors, an ion charge, for example. The effect of impurities strongly varies with the host materials used. For example, copper absorbs a factor of 30 and nickel a factor of 10 less energy in quartz than in glass; on the contrary, chromium and cobalt impurities absorb more energy in quartz.

Heavily absorbing centers form ions of the hydroxyl group, OH, which are inevitably present in the fiber material. The scattering loss varies with the amount of fabrication defects such as bubbles and crystalline inclusions. If the improved technology eliminates these defects, the only loss that is likely to occur is the loss due to an inhomogeneous distribution of the components of the material over its volume. This loss determines the theoretical lower limit for fiber guides. It is near 0.7 dB/km at $\lambda = 1$ μm for quartz and somewhat higher for multicomponent glasses.

Optical cable fibers include additional elements which impart elasticity and strength to cables. Structurally, fiber optic guides are rather diverse, but have a number of features in common (Fig. 4.23).

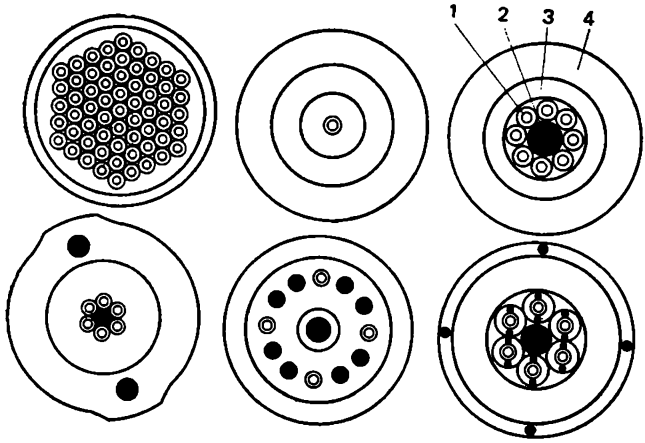


Fig. 4.23. Optical cables designed for various applications
(*1*) fiber; (*2*) strengthening element; (*3*) inner coating; (*4*) outer coating

Each fiber is coated with thin protective film to exclude stray pickup between the fibers stranded in a bundle. The fiber bundle is given a common coating over its entire length that combines elasticity with a certain degree of rigidity. The outer braiding gives the cable mechanical strength and protects it against external influences in operating conditions. All protective claddings are made from polymer materials.

4.7. Holographic Data Systems

4.7.1. Principle of holography. Looking at a photograph, try to see the scene behind the image of the object as you change the angle of view. The attempt will certainly be a failure because the photograph is a two-dimensional image picked up from a certain point of observation. Holography offers this opportunity. Object waves recorded on the photoplate and duly restored can give a 3-d image of the object as it actually appears in space, its color, and side views on changing the direction of the reference beam. Holography is the method of recording and subsequent reproduction of object waves. To record a 3-d image of the object of interest, we need to gain information on both the amplitude and the phase of the spatial wave reflected from the object. A change in the phase of the object wave can be written in the hologram only if this wave is made to interfere with another wave coherent with the first (see Subsec. 1.1.3).

So, holography relies on the principle of interference of an object wave with a reference wave of the same coherence. The amplitude of the resultant wave, i.e. the interference pattern produced by the

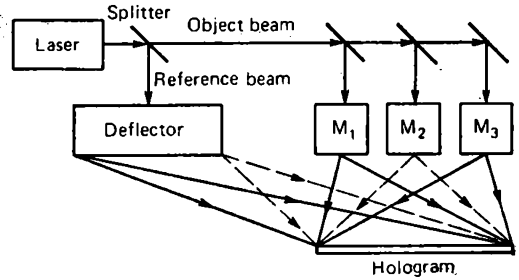


Fig. 4.24. Block diagram of a holographic memory

two superposed coherent waves and recorded on the photoplate, carries information on both the amplitude and phase of the object wave. The pattern so obtained is a hologram. To reproduce the object image, we must illuminate the hologram by the reference wave. The diffraction of this wave on the interference pattern makes it possible to restore the original object wave. Coherent optical signals thus enable us to process multidimensional information on images or information transmitted along many parallel channels.

4.7.2. Holographic memory. A simple scheme of data recording in a holographic memory is shown in Fig. 4.24. The laser beam passes through a splitter to form two beams, an information-bearing (object) beam, or write beam, and reference beam. The write beam travels through the system of modulators M_1, M_2, . . ., and falls on the photoplate. The reference beam impinges on a deflector and also arrives at the photoplate. The figure demonstrates two paths of recording shown by full and dash lines. In the first path modulators M_1 and M_2 are open and the deflector routes the reference beam from its left part. In the second path, a modulator M_2 is open and the reference beam progresses from the right part of the deflector. In each case the direction of the reference beam going from the deflector is different. It is thus possible to record a host of various two-dimensional sets of luminous point images on the same hologram. These point images are carriers of definite digital information. This information can be read out by exposing the hologram to a beam that is exactly the same as the reference beam. In the described example, information is entered into the entire photoplate, and new information is put in by changing the reference beam direction. A hologram divided into small regions, each serving as an elemental hologram, affords a higher recording density. In this case the elemental holograms form what is called a hologram matrix. The deflector enables the write beam to scan across the matrix and record a certain body of information in each element of the matrix.

Thus, the holographic memory is a storage matrix where each

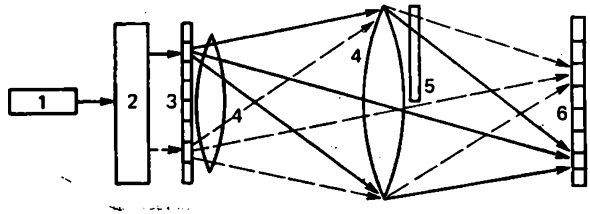


Fig. 4.25. Diagram of a holographic recording device

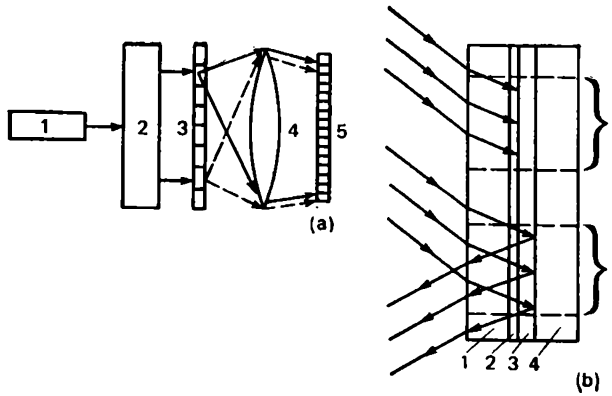


Fig. 4.26. Diagram of a holographic readout device
(*a*) separate arrangement of storage matrix; (*b*) composite arrangement

matrix element stores a two-dimensional interference pattern of two light waves, the reference wave and the write wave which is the information carrier modulated by the information signal.

4.7.3. Holographic recording and readout devices. Fig. 4.25 illustrates the block diagram of a recording device which includes a laser *1*, deflector *2*, auxiliary hologram matrix *3*, collecting lens *4*, space-time modulator *5*, and storage hologram matrix *6*. Assume the deflector *2* routes the laser beam (full line) to the matrix *3*. The diffraction of light causes two waves to appear: one is the reference wave directed to the memory matrix *6* by lens *4* and the other is the spatially modulated write wave also routed to the matrix *6*. If the deflector plays the beam on another element of the auxiliary matrix, the reference and write waves will travel in other directions (dash line) to enable another element of the storage matrix to receive information. The described device is used to insert information received from a computer into an optical memory (hologram matrix). The computer information arrives at the space-time modulator in the form of electric signals modulating the laser beam.

The block diagram of a holographic readout device is shown in Fig. 4.26*a*. It contains a laser *1*, deflector *2*, lens system *4*, storage

matrix *3*, and photodetector (PD) matrix *5*. In the readout cycle, the deflector *2* directs the beam to a certain holographic element of the matrix *3* to reproduce the image. The lens system now focuses the optical signal of the image on to the photodetector matrix *5* to convert it to a train of electric pulses.

Modern holographic systems commonly use a composite device called a 'latrix' that combines the function of the data input matrix with that of the PD matrix. Fig. 4.26*b* illustrates the schematic outline of a 'latrix' which includes a glass plate *1*, transparent conductor *2*, liquid crystal *3*, and silicon plate *4*. The silicon plate carries a matrix of optical switches, each consisting of a PD and a flip-flop. Such a switch scatters light in the on state, and reflects it in the off state. The control voltage applied to a certain portion of the liquid crystal enables the switch to be on, otherwise it is off. In the write cycle, control signals secure a certain combination of on and off switches to enable spatial modulation of the intelligence wave. In the read cycle, all switches are on and the beam freely passes through the layer of the liquid crystal and comes to PDs. Each photodetector drives a certain flip-flop which stores information on the state of the PD.

Chapter 5

Fabrication of Optoelectronic Devices

5.1. Fabrication Processes

To get an insight into the procedures of manufacture of optoelectronic devices, we should be familiar with the basic fabrication processes and techniques involved in the manufacturing cycles. Let us consider them.

5.1.1. Semiconductor wafer production. Single crystal semiconductor ingots are generally grown from the melt, for example, by the Czochralski technique of crystal pulling (Fig. 5.1). A single crystal of, say, silicon serves as a seed. The pulling rod with the seed crystal at its end is lowered into contact with the melt and then slowly raised and rotated simultaneously. The liquid column suspended from the seed and gradually pulled up solidifies into a single crystal ingot typically 80 mm in diameter and tens of centimeters in length. The maximum diameters and lengths of ingots available today can be over 120 mm and 1000-1500 mm respectively.

The ingots so obtained are cut into thin blanks 0.4 or 0.5 mm thick with a diamond saw or disk. Each ingot must be correctly located

with respect to the saw blade and rigidly fixed to obtain the slices, or wafers, of the desired crystallographic orientation.

The blanks are then repeatedly lapped flat and polished to remove the mechanical defects left after sawing and also to bring the wafers to the specified thickness with the faces parallel to each other. Lapping is done on special turntables using micropowder suspensions as an abrasive material. In each successive cycle of lapping the suspension of a lower grain size is chosen, down to 1 or 2 μm.

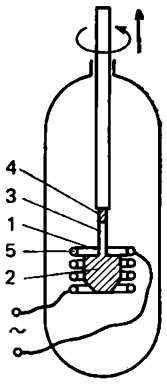


Fig. 5.1. Czochralski technique of single crystal pulling (*1*) crucible; (*2*) melt; (*3*) ingot; (*4*) seed crystal; (*5*) induction coil

After lapping the wafers need further treatment to remove two disturbed layers on the surface, namely, a mechanically distorted upper layer a few micrometers thick and a yet thinner sublayer having microdistortions in the crystal lattice and residual stresses caused during lapping.

Finishing polish removes the two disturbed layers. The surface unevenness is brought down to hundredths of a micrometer by mechanical or chemical polishing. Mechanical polishing levels out the surface with the aid of yet finer-grained micropowders and suspensions. Chemical polishing is the process of etching of the surface layer with suitable solvents. The etch actively dissolves protrusions but has no time to disturb the basic material, thereby making the surface smooth.

An important preparatory operation is to clean the surface of wafers from organic contaminants, greases in particular. Degreasing is done at elevated temperatures using such solvents as acetone, trichloroethylene, and carbon tetrachloride.

After each operation of etching, cleaning, degreasing, etc., the wafers are rinsed in deionized water to render the next mechanical operation more effective. Deionization is carried out in special installations by passing the distilled water through granular resins which 'bind' ions dissolved in water. The degree of deionization is defined in terms of water resistivity which usually ranges between 0.01 and 0.10 MΩ m and above; the resistivity of distilled water is equal to about 8 kΩ m.

5.1.2. Epitaxy. This is the process of oriented growth of single crystal layers on a substrate from the vapor (gas) or liquid phase. In this process the crystallographic orientation of the layer grown repeats that of the substrate material. The epi process yields a thin

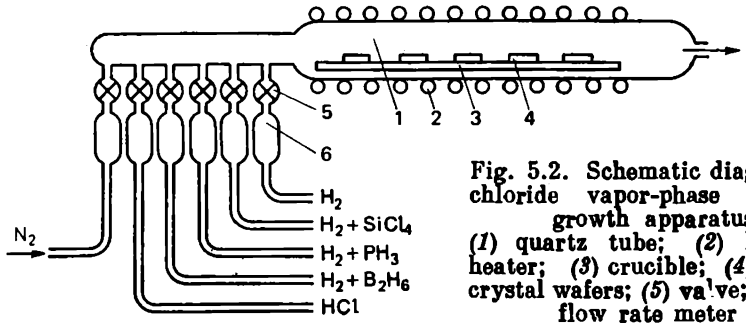


Fig. 5.2. Schematic diagram of chloride vapor-phase epitaxy growth apparatus (*1*) quartz tube; (*2*) hf coil heater; (*3*) crucible; (*4*) single crystal wafers; (*5*) valve; (*6*) gas flow rate meter

active semiconductor layer on a rather thick passive substrate which serves as a support for active elements.

Consider the process of epitaxial growth on, say, silicon wafers (Fig. 5.2). A crucible loaded with single crystal silicon wafers is placed inside a quartz tube. The crucible is inductively heated up to 1 200°C by a hf coil heater. Then hydrogen charged with vapors of silicon tetrachloride $SiCl_4$ is passed through the tube to initiate the chemical reaction on the wafer surfaces:

$$SiCl_4 + 2H_2 = Si + 4HCl \uparrow$$

A layer of pure silicon gradually grows on the wafers which play the role of substrates, the vapors of HCl being carried away by the stream of hydrogen. The deposited epi layer continues the single crystal structure of the substrate. Temperature control causes the reaction of silicon deposition to proceed only on the substrate.

The described process is the process of chloride vapor-phase epitaxial (VPE) growth where hydrogen that transfers reactants to the reaction zone is known as a carrier gas.

If the carrier gas transfers the vapors of diborane B_2H_6 added to $SiCl_4$, the acceptor atoms of boron will diffuse into the growing layer of silicon. The epi layer of the intrinsic conductivity will turn to the *p*-type layer. The additions of phosphine PH_3 will yield the *n*-type epi layer.

In the setup of Fig. 5.2 provision is made to perform additional operations such as blow-through of the tube with nitrogen and shallow etching of silicon wafers in HCl vapors to clean the wafer surface.

Epitaxy enables the oriented growth of *p* and *n* layers in the desired resistivity range on the substrate of any type. Fig. 5.3 illustrates the epi structure of p^+p, nn^+, p^+n, and np types.

The epi process can be homoepitaxial and heteroepitaxial depending on the materials of the epi layer and the substrate. In homoepitaxy the deposited layer and substrate, are of the same chemical nature, for example, the layer of silicon deposited on the silicon substrate. In heteroepitaxy, the layer and substrate differ in chem-

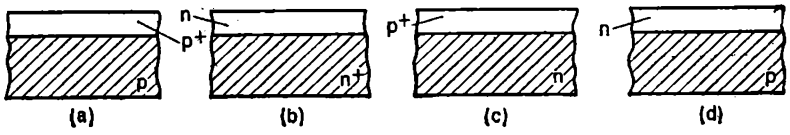


Fig. 5.3. Epitaxial structures
(*a*) p^+p type; (*b*) nn^+ type; (*c*) p^+n type; (*d*) np type

ical composition. Of course, the basic parameters of the crystal lattices of both materials, for example, a silicon film and a sapphire substrate, must be the same.

The boundary between the epi layer and substrate cannot be made ideally abrupt since impurities partially diffuse from one layer into the other in the course of epitaxial growth. That is why the epi process cannot produce very thin films less than 1 μm thick and multilayer structures.

Liquid phase epitaxial (LPE) growth is also popular in optoelectronics, along with vapor phase epitaxy. By this technique, single crystals are grown from the liquid phase, which is the solution containing requisite components.

5.1.3. Oxidation. Silicon is the most widely used material in semiconductor electronics. Silicon oxidation is one of the basic processes in modern IC technology. The film of silicon dioxide SiO_2 obtained by this process performs two important functions: (1) protects (passivates) the semiconductor surface, in particular vertical *pn* junction regions coming out to the surface, against external influences; and (2) serves as a mask defining windows for local diffusion of dopants.

Silicon oxidation generally requires high temperatures, from 1 000 to 1 200°C, therefore this process of SiO_2 film growth is known as thermal oxidation. SiO_2 films are grown in the atmosphere of oxygen (dry oxidation) or in the mixture of oxygen and water vapors (steam oxidation). A typical oxidizing furnace used for the purpose consists of a quartz tube accommodating a boat with silicon wafers and a heater to maintain the required temperature in the reaction zone while forcing oxygen or water vapors through the tube. Oxygen reacts with silicon to yield a SiO_2 film of the amorphous structure.

One of the important problems involved in the oxidation process is to ensure the film homogeneity. Imperfection in the surface layer of a wafer, insufficient purity of the reagents, and inadequately controlled oxidation conditions reduce the yield of acceptable wafers. The widespread types of defect are micropores, macropores, and even pinholes, especially in thin oxide films. The quality of SiO_2 films generally increases in the conditions of dry oxidation and at decreased oxidation temperatures. But the rate of dry oxidation is

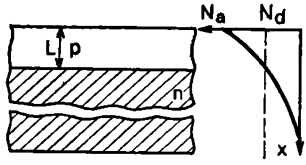


Fig. 5.4. Diffused *pn* junction

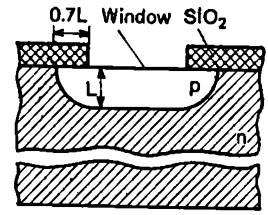


Fig. 5.5. Local impurity diffusion

lower than that of high-temperature oxidation in water vapors. Therefore, thick films are grown in the atmosphere of dry oxygen and wet oxygen alternately: dry oxidation reduces the number of defects in films and wet oxidation cuts down the total time required for the entire process.

5.1.4. Doping. This is the process of introduction of impurities into the starting material of a semiconductor wafer or an epi layer. Semiconductor technology employs two methods of doping, diffusion and ion implantation, of which the first is most popular.

The diffusion process gives a layer of nonuniformly distributed impurities. Figs. 5.4 and 5.5 illustrate a diffused p layer obtained by doping the material with acceptor impurities and also show the impurity distribution $N_a(x)$ in depth L of the layer.

The diffusion process can be run repeatedly to produce multilayer structures with several *pn* junctions. The first diffusion produces, for example, a p layer after doping the starting n-type wafer with acceptor impurities. The second diffusion can form an n layer in the p layer if a donor impurity is driven into the p region to a smaller depth. The result is a three-layer structure. The donor concentration introduced during the second diffusion stage must obviously be higher than the acceptor concentration provided in the first stage, otherwise the conductivity type will remain the same. However, the impurity concentration in any starting material has an upper limit determined by the critical parameter called solid solubility N_s of an impurity. This parameter is temperature dependent. It reaches its maximum value, N_{max}, at certain temperatures and then begins to decrease.

Dopant	As	P	B	Sb
N_{max}, cm^{-3}	20×10^{20}	13×10^{20}	5×10^{20}	0.6×10^{20}
	1150°C	1150°C	1200°C	1300°C

In the last stage of multiple diffusion, the solid solubility of the impurity must certainly be the highest over N_s of the impurities introduced in the preceding stages. However, the narrow range of

appropriate impurity materials as regards their values of N_s generally limits the diffusion procedure to three sequential steps.

Diffusion can be overall to produce a diffused layer over the entire wafer surface and local to form layers in definite portions of the wafer through the windows opened in, say, a SiO_2 layer which serves as a mask. In local diffusion the impurity penetrates vertically into the wafer bulk and spreads over under the mask. The mask protects the *pn* junction portions extending outward to the surface. The lateral diffusion depth is typically 0.7 L.

The sources of impurities, called dopants or diffusants, are chemical compounds in various aggregate states. These can be gases (B_2H_6, PH_3), liquids (BBr_3, PCl_3) and solids (B_2O_3, P_2O_5).

The diffusion process is carried out in single-zone or double-zone diffusion furnaces. Single-zone furnaces largely use liquid or gaseous diffusion sources. The reaction proceeds in the atmosphere of oxygen added to the carrier gas. Oxygen reacts with silicon to form silicon dioxide SiO_2. In the presence of the diffusant boron or phosphorus the reaction yields a borosilicate or phosphosilicate glass. These glasses are in the liquid state at above 1 000°C. The glass solidifies to give a thin film which protects the silicon surface.

The diffusion process in semiconductors obeys Fick's laws. Fick's first law relates the impurity flow density y to the concentration gradient of the impurity:

$$y = -D\ (dN/dx) \tag{5.1}$$

where D is the diffusion constant.

Fick's second law defines the rate of accumulation of impurity atoms, i.e. a change of the concentration in time:

$$\partial N/\partial t = D\ (\partial^2 N/\partial x^2) \tag{5.2}$$

The diffusion process heavily depends on temperature. Fig. 5.6 illustrates the plots of diffusion constants versus temperature for some materials used as dopants of silicon in the manufacture of optoelectronic devices. As is clear from the graphs, the temperature dependence is very abrupt: with a change of temperature by 100°C, the diffusion constant varies by a factor of 10; the rate of change of D per degree Celsius is 2.5%. This points to the need of maintaining extremely stable temperatures in the diffusion furnaces; a change in temperature should not exceed $\pm$(0.25-0.5°C), or tenths of a percent.

Another method of doping a starting material is ion implantation by bombarding the semiconductor wafer with impurity ions at an energy enough to enable the ions to penetrate deep into the wafer bulk. Ion implantation has a number of features which distinguish it from diffusion:

(1) the depth of ion penetration is much lower than the diffusion depth and does not exceed 0.4 μm;

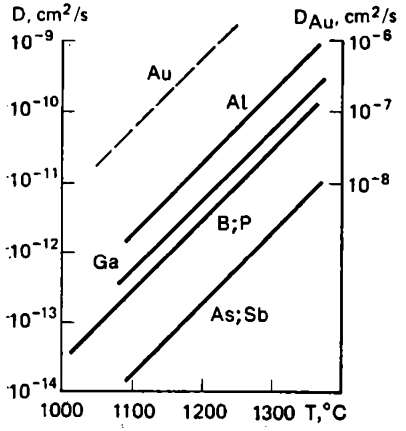


Fig. 5.6. Diffusion constant versus temperature for Au (right scale) and other elements (left scale)

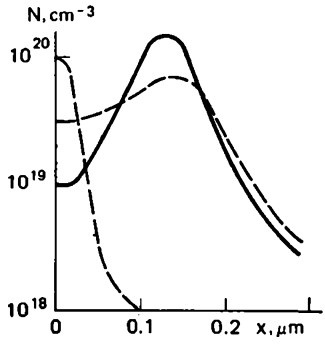


Fig. 5.7. Impurity distribution in a semiconductor during ion implantation

(2) the impurity distribution displays a peak, as illustrated in Fig. 5.7, and much differs from that found in the diffusion process;

(3) doping through the windows proceeds without lateral spreading specific to diffusion;

(4) formation of thin single layers is the main purpose of the process since multiple doping involves manufacturing difficulties;

(5) low temperature of doping is the main advantage of the process.

Special setups are used to generate the accelerated flow of impurity ions. The depth of impurity penetration depends on the energy and mass of ions. The thickness of the implanted layer is proportional to the ion energy. But the increased energy of ions gives rise to radiation defects in the crystal and impairs its electrophysical parameters. The ion energy ranges from 10 to 1 500 keV, which limits the implanted layer thickness to 0.1-0.4 μm.

Since the ion beam is commonly much smaller in cross section than the wafer subjected to doping, the setup has a special beam deflector to scan the beam line by line across the entire wafer surface if need be.

After completing the implantation procedure, wafers are annealed at 500-600°C to reduce radiation-induced defects.

5.1.5. Photolithography. The photolitho technique, also called photomasking and photoengraving, is the process intended to form the oxide mask of the desired pattern and size on a dielectric or metal plate using for the purpose photomasks (Fig. 5.8). The photomask is a thick glass plate one side of which has a thin nontransparent coat with the pattern of transparent openings. These openings or

pattern elements can be equal in size to the appropriate circuit elements of photoelectronic devices. Since the devices are produced by the batch technique, the photomask contains a large number of pattern elements.

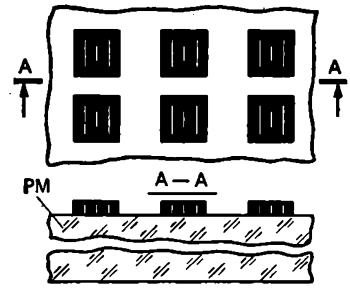


Fig. 5.8. Photomask

Photolithography uses photoresists to open windows in the SiO_2 mask covering the wafer surface. These are UV light-sensitive photoemulsions which come in negative-acting and positive-acting types. The negative photoresist polymerizes under light and becomes stable to etchants such as acidic and alkaline solutions; after its selective exposure to light and development, the photoresist will have the exposed portions stable to the etch. On the contrary, in positive photoresists the light breaks polymer chains, so that the etch will destroy the exposed portions leaving the unexposed portions intact.

The photolitho process for opening windows in the SiO_2 films consists of several steps (Fig. 5.9). A drop of photoresist PR is allowed to fall on the SiO_2 film and the wafer is rotated to spread the photoresist over its surface in a film of about 1 μm thick. The photomask (PM) with its pattern facing the photoresist is now placed over the wafer and exposed to the light of a quartz lamp (Fig. 5.9*a*). If the photoresist used is of the positive-acting type, the photomask is also positive: the transparent portions on the photomask correspond to the windows subsequently opened in the photoresist layer. The negative photoresist calls for a negative photomask: the photoresist will have openings in the areas corresponding to the nontransparent portions on the photomask. Subsequent to light exposure, the photomask is taken off.

After its exposure and development, the photoresist becomes the mask and bears the replica of the pattern on the photomask. Next an etch neutral to the starting semiconductor is applied to remove the

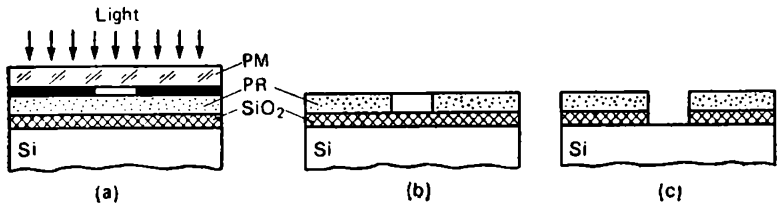


Fig. 5.9. Photomasking steps to open windows in the SiO_2 film (*a*) photoresist deposition, photomask placing, and exposure; (*b*) photoresist development; (*c*) SiO_2 film etching

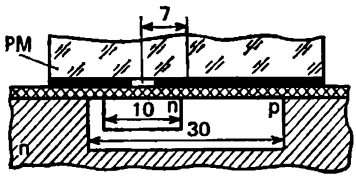


Fig. 5.10. Alignment of the photomask pattern with the photoetched pattern (dimensions in μm)

oxide layer as far as the silicon layer through the windows in the photoresist mask (Fig. 5.9*b*, *c*). In this step of etching, we have ransferred the pattern on the photoresist film to the oxide layer. The final step is to etch away the remaining photoresist leaving intact the oxide mask with windows.

In the manufacturing cycles involving the fabrication of optoelectronic devices, photomasks are repeatedly used to form, say, the base, emitter, and ohmic contacts. Assume the first masking step and subsequent diffusion have provided a *p* layer 30 μm wide in the *n*-type wafer. In the next similar procedure, an *n* layer 10 μm wide need be driven into the *p* layer with its centerline shifted 7 μm away from the *p* layer centerline (Fig. 5.10). For this the pattern of the second photomask and the pattern (boundaries) of the *p* layer must register to within 1 μm. The multiple photomasking procedure including 5 to 7 operations calls for tight alignment tolerances, down to fractions of a micrometer. The alignment is done under a microscope with the aid of registration marks (positioning aids) provided on the subsequent photomask and the oxide layer. One manufacturing cycle requires a set of photomasks, each destined for the particular masking operation. All photomasks of the set are compatible, i.e. they have their patterns in register as the superposed marks are brought into coincidence.

Contact photomasking described above suffers from certain inherent drawbacks which make themselves felt more strongly as semiconductor electronics makes further advances. These drawbacks include: (1) the limit of resolution, i.e. the limit to the fineness of detail of photomask pattern elements which cannot be brought down below 1 μm; (2) mechanical contact of the photomask with the substrate, which causes pattern distortions and photomask wear; (3) difficulties involved in mask alignment.

Specialists believe that projection photolithography and electron beam lithography which are gradually gaining ground in the domain of optoelectronic technology will do away with the above limitations and difficulties.

5.1.6. Assembling. After completing the diffusion and metallization stages, the wafer incorporating many ICs is ready for scribing, or dicing. Vertical and horizontal grooves are cut with a diamond scribe along the boundaries of individual structures (Fig. 5.11*a*). The wafer is then placed on a soft rubber pad to break away the chips along the grooves (Fig. 5.11*b*). Sound chips are then mounted in packages.

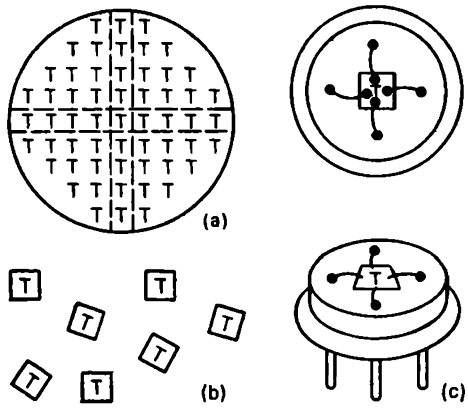


Fig. 5.11. Technique of scribing and separating the wafer and mounting the chip on the header

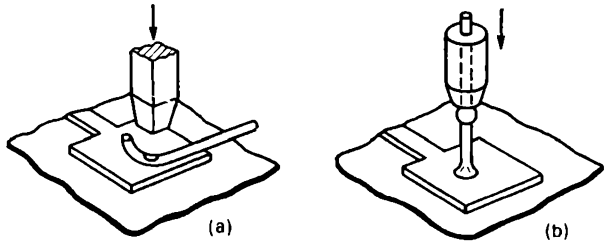


Fig. 5.12. Thermocompression bonding with a blunted wedge (*a*) and capillary tube (*b*)

The assembly of the device begins with what is called mounting the chip on the header (Fig. 5.11*c*). The chip is bonded or soldered in the middle of the header. The bonding pads of the chip are then connected to the lead-out frame of the header with jumpers which are fine aluminum or gold wires 25 to 50 μm thick.

Reliable connection between the bonding pads and the lead frame is usually made by thermocompression bonding. This technique requires a rather high pressure to enable good contact between the parts and fairly high temperatures, 200 to 300°C. This combination of pressure and temperature ensures mutual diffusion of atoms between the two parts and provides a solid weld. Fig. 5.12 illustrates typical examples of thermocompression bonding. Fig. 5.12*a* shows a wedge bonder which presses the wire against the pad, thus producing a firm stitch bond as it squeezes the wire and forces it along the pad. Fig. 5.12*b* shows a nailhead (ball) bonder, which is a capillary tube with a wire threaded inside. At the exit of the capillary the wire is cut off with a microtorch to form a liquid ball at the

wire end. The capillary is now brought in contact with the pad to spread over the ball and to produce a firm contact in the form of a nailhead. The bonder is then raised for a length of wire sufficient to connect it to a lead. The wire is again cut off with a flame.

Once mounted and connected to the lead frame, the chip need be encapsulated. This is done by hot or cold welding of the header to the cap to provide the sealed envelope. The purpose of sealing is to exclude the adverse effects of the environment on the chip, therefore this operation is performed in a vacuum or in a protective atmosphere such as dry nitrogen or argon. Some circuit elements do not require packaging. These are discrete add-on elements used in hybrid ICs.

5.2. Fabrication of LEDs

5.2.1. Production of single crystal semiconductor compounds. Light-emitting diodes, just like silicon diodes, are fabricated by planar technology involving the basic processes described in the preceding section. But the specific requirements placed on the source materials used for LED fabrication necessitate certain modifications in the technological steps. As mentioned earlier, these materials must be wide-gap semiconductors because the maximum energy of photons is proportional to the gap energy. However, the wider the bandgap of a material, the more difficult is the process of imparting the material a higher photoconductivity. So, the trade-off is to use semiconductors whose gap energy is not more than 3.5 eV.

The source materials suitable for the manufacture of LEDs are both direct-gap semiconductors, as required for injection lasers, and indirect-gap semiconductors where radiative recombination occurs through impurity centers. In deciding on the materials for devices, preference is given to semiconductors with a high mobility of free carriers because they let pass through larger currents. The choice of materials appreciably depends on the degree to which they can retain their electric and especially electroluminescent properties.

The materials most widely used today for the fabrication of LEDs in visible applications include gallium phosphide, silicon carbide and also the solid solutions of gallium arsenide-phosphide and aluminum-gallium arsenide. The most common materials for IR diodes are gallium arsenide and solid solutions of aluminum-gallium arsenide.

At room temperature the bandgaps for gallium phosphide, gallium arsenide, and silicon carbide are equal to 2.25, 1.45, and 2.95 eV respectively.

The bandgap in silicon carbide can vary with the lattice structure from 2.35 to 5.9 eV. Various modifications of silicon carbide crystals can apply for fabrication of LEDs with different spectral characteristics.

The use of solid solutions opens up even more possibilities for extending the emission spectra of LEDs. Changing the ratio between the two components of a solution enables the manufacture of a LED structure with emission spectrum peaks of 1.4 to 2.2 eV. Also, adequately varying the composition of the wafer in depth gives the structure that can absorb a less amount of light in its bulk.

The material selected for LED manufacture must also fill additional requirements along with the basic requirement for a large bandgap.

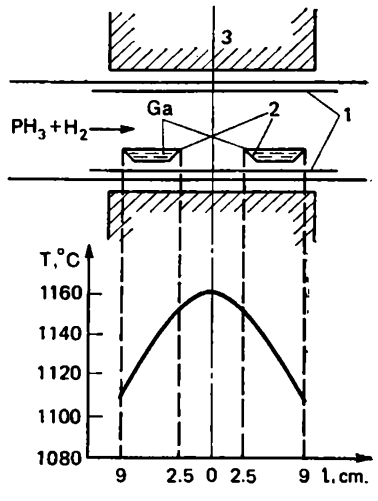


Fig. 5.13. Schematic diagram of low-pressure GaP synthesis apparatus *(1)* **inner tube;** *(2)* **boat with liquid gallium;** *(3)* **heater**

1. It must be free from cracks, voids, dislocations, and other defects which result in nonuniform emission and decrease the efficiency of radiative recombination.

2. The impurity concentration must be 10^{18} cm^{-3} at the most to provide for fairly high conductivity and thus low power dissipation. Higher concentrations commonly promote excessive light absorption in the material.

The technique for the production of single crystals of compound semiconductors is more complex than that used for growing single crystals of elemental semiconductors. For comparison, the synthesis and growth of GaAs or GaP single crystals commonly used for substrates can be taken as an example. In contrast to silicon, GaAs and GaP compounds have high pressures of arsenic and phosphorus vapors at the melting point. The synthesis and single crystal growth can be carried on separately or in a unified process, at low pressures when using gallium-enriched melts or at higher pressures in the case of stoichiometric melts.

Consider a few industrial processes of GaP and GaAs single crystal growth.

GaP synthesis. As mentioned above, GaP single crystals can be grown at low and high pressures. The schematic drawing of a low-pressure GaP synthesis setup is shown in Fig. 5.13. The internal tube *1* from pyrolitic boron nitride BN encloses a boat-like crucible *2* from BN for liquid Ga. Phosphine PH_3 dissolved in hydrogen is forced into a hot reaction tube, where the phosphine decomposes into phosphorus and hydrogen. If the phosphorus vapors interact with a

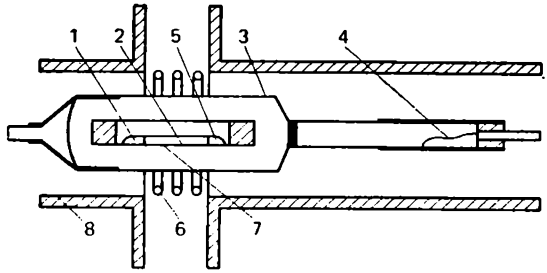


Fig. 5.14. Schematic cross-section of high-pressure GaP synthesis apparatus (*1*) gallium; (*2*) melt zone; (*3*) quartz vessel; (*4*) phosphorus; (*5*) GaP crystal; (*6*) induction heater; (*7*) graphite tube; (*8*) furnace

uniformly heated Ga surface layer, a solid crust of GaP may form, which will impede a further course of the reaction.

To eliminate this unwanted effect, gallium is held in the region of a high temperature gradient (near 10°C cm^{-1}), so that the rate of crystallization increases from the low-temperature to the high-temperature end of the boat. The solid phase first forms in the low-temperature region. This gives rise to the phosphorus concentration gradient in liquid gallium, which promotes phosphorus diffusion in the direction of the zone of higher temperatures. In typical conditions, 50 to 55% of PH_3 and gallium react with each other in the course of synthesis that lasts for 72 hours. The unreacted phosphorus condenses at the outlet and is burnt down. The unreacted metal left in GaP in the form of inclusions is cleared away with an acid. The process is simple and inexpensive, but utilizes the material inefficiently and proceeds slowly due to slow diffusion. The high-pressure GaP synthesis process obviates these limitations.

The diagrammatic sketch of a high-pressure apparatus is shown in Fig. 5.14. The apparatus is enclosed in a high-pressure stainless-steel chamber pressurized to 0.8-1.0 MPa. The quartz vessel contains a graphite boat for gallium and phosphorus. Phosphorus is heated to 510°C to raise the pressure to the required value, which is equal to 0.8 MPa. To keep the crucible from bursting, the chamber is held at the same pressure of nitrogen. Induction heating of the graphite boat produces a hot zone in gallium. The synthesis begins at the moment when one end of the gallium rod heats up to 1 460°C and the reaction proceeds as the quartz vessel moves through the hot zone at a speed of 1 cm/h. The chemical reaction yields a 40% solution of phosphorus in gallium. As the vessel is making way from left to right, the solution enriched in gallium solidifies into GaP. Inclusions of Ga can be canceled if the vessel is moved slowly. The amount of reacted gallium can reach 90 to 100%. A disadvantage of the process is that the graphite boat contaminates polycrystalline GaP with graphite, whose content can be as high as 0.1%. Replacement of

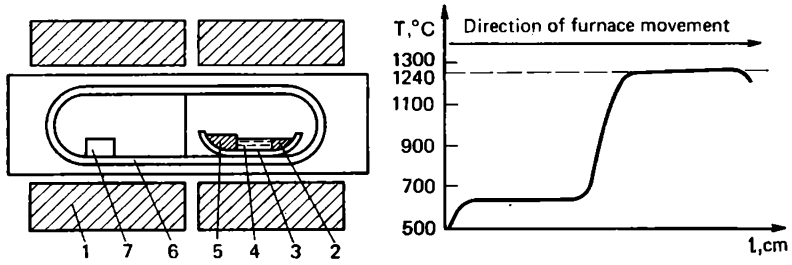


Fig. 5.15. Setup for GaAs crystal growth by the Bridgman technique (*1*) two-zone horizontal furnace; (*2*) gallium pieces; (*3*) boat; (*4*) melt; (*5*) GaAs crystal; (*6*) vessel; (*7*) crystalline arsenic

this boat by boron-nitride crucible obviates the disadvantage. This process requires more complex equipment to maintain high pressure, move the vessel, and ensure hf induction heating. But it utilizes more efficiently the source material, takes less time to accomplish the synthesis, and produces a denser polycrystalline material suitable as a source material for pulling single crystals from its melt under a flux.

Crystal growth from the melt. The Bridgman technique of horizontal growth is more economical and suitable for GaAs, and the technique of synthesis and diffusion in the solution for GaP. The latter technique gives the polycrystalline material of such a purity that it can be used as a material for substrates.

A typical arrangement for crystal growth by the Bridgman technique appears in Fig. 5.15. The boat is loaded with gallium, and an appropriate amount of arsenic is laid in the opposite end of an evacuated quartz vessel mounted in a two-zone horizontal furnace. The synthesis begins when one end of the vessel with As heats to 620°C, and its other end with Ga to 1240°C. As the furnace moves on, the crystal grows in the direction of the temperature gradient along the boat from the hot region (1240°C) to the cold region (620°C). The nucleation starts at one end of the melt and continues along it. It is also possible to use a seed to initiate nucleation. This technique readily produces single crystals shaped in the form of a hemisphere 4 or 5 cm in diameter and oriented in the ⟨111⟩ direction. The dislocation density varies from 10^2 to 10^4 cm^{-2}.

The Czochralski technique of crystal pulling under the flux. This technique is more suitable than other methods for industrial production of large-diameter GaP crystals and GaAs ingots. Fig. 5.16 displays the schematic cross-section of a high-pressure crystal-pulling apparatus. The stoichiometric GaP melt contained in the SiO_2 crucible is covered by the molten boron oxide B_2O_3. The chamber is filled with an inert gas at a pressure of 5 MPa that is higher than the

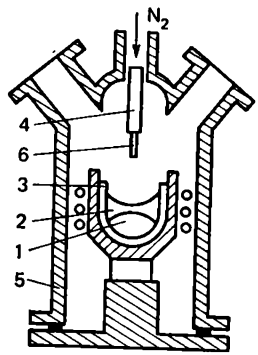


Fig. 5.16. Crystal pulling apparatus
(1) GaP melt; *(2)* B_2O_3 melt; *(3)* SiO_2 crucible; *(4)* pulling rod; *(5)* stainless steel jacket; *(6)* seed crystal

pressure of phosphorus vapors (3.5 MPa) above the GaP melt. The volatile component P of the crystal is thus held enclosed under the layer of the molten flux B_2O_3. The flux keeps the phosphorus vapors from escaping into the chamber, which otherwise would fill it up at a pressure above 3.2 MPa and heat the stainless steel jacket over 700°C. The apparatus operating at such a temperature and pressure and in the aggressive medium would require special sealing. The flux enables the chamber walls to be kept at room temperature, which makes for a simpler design of the apparatus. Nevertheless, the chamber walls of this setup stand up to high pressures and, besides, the process requires preliminarily synthesized GaP for the melt. Another disadvantage is that control over the crystal growth here is more difficult than is the case for a low-pressure setup. The process yields a good material for substrates in which active layers are grown by liquid epitaxy. But this material is not adequate enough for active light-emitting junction regions because the grown crystals have nonradiative recombination centers.

The technique of crystal growth under the flux is being rapidly improved. The efforts are taken to increase the capacity of pulling devices for GaAs and GaP and also to improve the degree of purity of grown crystals. This technique is most applicable for the manufacture of photoemitter substrates.

5.2.2. Fabrication of LED pn junctions by diffusion. The LED *pn* junction determines the quantum efficiency of its active region and the efficiency of light extraction from the semiconductor material. LED junctions in group $A^{III}B^{V}$ semiconductors can be grown by diffusion and epitaxy from the gas and liquid phases.The performance of semiconductor emitters depends not only on the techniques of *pn* junction fabrication but also on the composition and type of substrate material. For example, GaP is transparent to light generated in the *pn* junction, while GaAs is not. That is why diodes grown on GaP substrates have four times the efficiency of diodes accommodated on GaAs substrates.

Figure 5.17 illustrates the emission characteristics of LEDs, which can be of avail in deciding on the choice of these devices for particular applications. The LED from gallium phosphide doped with zinc and oxygen is the most efficient red lamp at low current densities. At high current densities, $GaAs_{1-x}P_x$ devices are most approp-

riate. Devices from nitrogen-doped GaP obtained by liquid phase epitaxy are the most efficient yellow-green light sources in the entire range of current densities.

Some features of diffusion processes. Diffusion and vapor phase epitaxy (VPE) are well developed techniques used to fabricate most of the discrete semiconductor devices and silicon ICs. They are also the basic methods for the fabrication of LEDs, though LED *pn* junctions grown by liquid phase epitaxy (LPE) are the most efficient emitters. In distinction from diffusion into silicon, diffusion into compound semiconductors takes a more complex course that complicates control over the process. The features peculiar to this type of diffusion are given below.

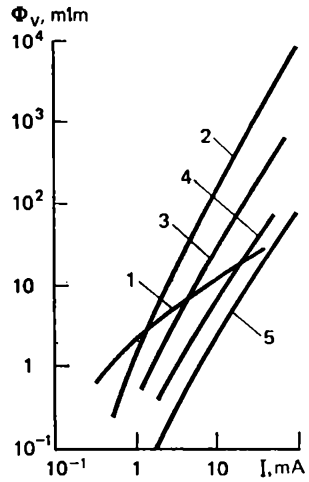


Fig. 5.17. LED emission characteristics
(1) red GaP LED; *(2)* yellow-green GaP LED; *(3)* orange-red GaAsP LED on GaP substrate; *(4)* yellow GaAsP LED on GaP substrate; *(5)* red GaAsP LED on GaAs substrate (*pn* junction area is 0.32 by 0.32 mm)

1. The solid solubility of impurities in the GaAs compound is describable in the framework of ternary systems. In the case of silicon, we need to consider only binary systems.

2. GaAs decomposes under heat to yield the liquid phase rich in gallium and also arsenide vapors. At a certain temperature the composition of the vapor phase is the same as that of the solid phase. This decomposition temperature for GaAs is equal to 637°C. Since electrically active impurities are commonly substituting agents, the vapor pressure of volatile components As or P strongly affects the quality of a diffused junction.

3. Diffusion constants D for acceptors are far higher than D for donors in semiconductor compounds. The factors affecting the D value at a specified temperature are the surface impurity concentration, position of the Fermi level, and vapor pressure of the volatile component. In order for a diffusant to be chosen, it is sufficient to compare the typical values of D for the commonly used acceptors and donors, namely, for Zn and S in GaAs at 1000°C and at surface concentrations in the range of 10^{19} cm^{-3}. The values of D for Zn and S are 10^{-9}-10^{-8} cm^2/s and 10^{-12}-10^{-11} cm^2/s respectively. Because of this high difference in D values, only the diffusion of an acceptor impurity, zinc as a rule, is made at 700-1000°C to form *pn* junctions in all LEDs manufactured on an industrial scale.

4. In the manufacture of silicon devices, the silicon dioxide SiO_2

serves as a mask. For $A^{III}B^{V}$ semiconductors, there are no readily obtainable oxide masks from the elements of these compounds. The masks used for Zn diffusion are analogous to those applied in the fabrication of silicon devices. These are SiO_2, Si_3N_4, and Al_2O_3 commonly employed in combination with one another, for example, SiO_2 with Si_3N_4. The silicon dioxide deposited over the Si_3N_4 layer protects this layer against the action of the etchant such as phosphoric acid; SiO_2 deposited below Si_3N_4 keeps the hot phosphoric acid away from the semiconductor compound surface and protects it against thermal decomposition during diffusion. The SiO_2 film deposited on top of the Al_2O_3 film weakens the mechanical stress at the oxide-semiconductor interface since it induces the compression stress and counterbalances the pulling stress caused by the Al_2O_3 film.

Plasma etching can markedly improve the masking action of oxides. Plasma CF_4 removes SiO_2 or Si_3N_4 leaving intact the semiconductor surface. Consequently, the Si_3N_4 layer can be grown on top of the photoresist mask and then appropriately removed by plasma etching. Following this, the photoresist mask can be etched away by replacing the medium in the chamber containing plasma (changing CF_4 for oxygen).

In the diffusion process used to form LED *pn* junctions, the parameters and properties destined for control are the junction depth, inclusions tending to impair crystal homogeneity, optical properties of the diffusion layer, and nonradiative centers arising in the course of diffusion. These parameters can be controlled by properly selecting a diffusion source, temperature, diffusion rate, and semiconductor material. As mentioned earlier, the diffusant used is commonly zinc. Diffusion occurs by the interstitial mechanism with substitution for host atoms.

The composition of substrates has an effect on the diffusion constant and junction depth which monotonically varies with composition. The junction depth for most practical ternary compound structures ranges from 5 to 15 μm. The diffusion source for $GaAs_{1-x}P_x$ is usually $ZnAs_2$ heated to 800°C. The diffusion process takes 45-60 min to drive in the diffusion to a depth x_j of 5 μm. At 850°C, the same source provides deeper junctions, up to 10 μm. The totally vaporizing source ZnP_2 heated to 900°C and diffused for 16 hours can give *pn* junctions in GaP 15 μm deep. The factors promoting the homogeneity of diffused junctions are the improved quality of substrates that have a low dislocation density and increased pressure of arsenic and phosphorus vapors above the substrate when using a totally vaporizing diffusion source such as $ZnAs_2$ or ZnP_2 enclosed in a sealed vessel.

The diffusion profile becomes uniform at a lower diffusion rate, i.e. at a smaller junction depth (Fig. 5.18). Increasing the arsenic

pressure gives rise to Ga vacancies which, in turn, contribute to the slow diffusion along lattice sites (the interstitial and the lattice diffusion proceed simultaneously), with the result that the diffusivity of Zn decreases. A decreased diffusion rate makes it possible to equalize the bulk vacancy concentration, so the diffusion front encounters uniformly distributed Ga vacancies.

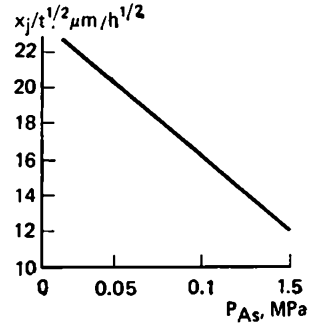


Fig. 5.18. Time-related *pn* junction depth in $GaAs_{1-x}P_x$ versus As_4 vapor pressure in Zn diffusion (in log-log scale)

Some features of diffused LEDs. The external quantum efficiency of the best diffused LEDs fabricated from direct-gap materials is about half as high as that of the best LEDs produced by the LPE technique and merely one third or one fourth that of red GaP LEDs. For yellow and green GaP diodes, the advantages of the LPE technique gradually fall into a decline as the low-temperature diffusion technology (620-650°C) is gaining ground. Initially, the diffusion process harnessed for the manufacture of red and green devices was run in the temperature range 800-1000°C. The quantum efficiency of green diodes was typically 0.03%. The diodes formed by VPE with the postgrowth diffusion of zinc into nitrogen-doped GaP at 620-650°C have half the quantum efficiency of the GaP diodes grown by the LPE technique. The former make efficient yellow-orange devices, whereas the latter only emit green to yellow-green light. This is because the material grown from the vapor phase lends itself to stronger doping with nitrogen.

The limitations specific to diffusion processes rest on the following.

1. In the materials formed by diffusion a fraction of diffused zinc atoms are electrically neutral. Approximately 20% of zinc atoms are neutral in GaAs compounds doped at 750°C and 50% in GaP compounds doped at 900°C. The difference between the total Zn concentration and active Zn—O concentration may be attributed to the deposition of Zn in defected regions of the crystal, which tends to increase the number of nonradiative centers.

2. The zinc concentration decreases with distance into the bulk, with the result that the light absorption in GaP increases and the light output goes down.

3. Radiative Zn—O centers form more readily during LPE growth of red GaP emitters even at low-temperature annealing.

The internal quantum efficiency η_e of diodes produced by double LPE and by diffusion is equal to 9% and 3.5% respectively. A high

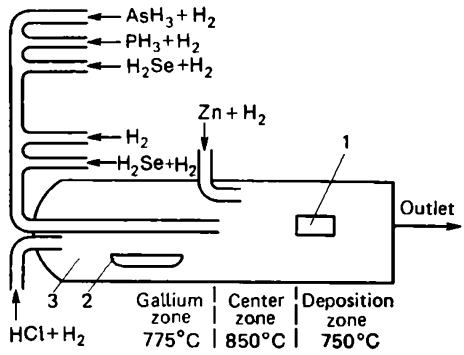


Fig. 5.19. Vapor phase epitaxy growth apparatus
(1) substrate; *(2)* liquid gallium; *(3)* quartz tube

Zn concentration in the region of doping is responsible for a lower light output efficiency of diffused diodes in comparison with the efficiency of epitaxial diodes. The total external quantum efficiency η of double-LPE grown diodes is 5%, while that of diffused diodes is merely 1.4%.

So, diffused LEDs have about one third of the quantum efficiency of LEDs grown by LPE. The advantage of the diffusion technique is that it can produce *pn* junctions in limited regions on the substrates formed by either the LPE or the VPE technique.

5.2.3. PN junction formation by vapor phase epitaxy (VPE). This technique, originally developed to produce silicon devices, now enjoys wide uses for the fabrication of silicon ICs and also $A^{III}B^{V}$ semiconductor compounds. There are many variations of the VPE technique for deposition of $A^{III}B^{V}$ compounds, for example, the sealed-capsule method and the open-tube method. Consider the technique of GaAs growth during transport and absorption of reactants in the open flow system. The chemical reactions involved in GaAs growth, which are analogous to those occurring in GaP preparation, include the reactions of synthesis and transport

$$2AsH_3(g) \rightleftarrows \frac{1}{2}As_4 + 3H_2(g)$$

$$2HCl(g) + 2Ga(l) \rightleftarrows 2GaCl(g) + H_2(g)$$

and epitaxy (deposition)

$$2GaCl(g) + \frac{1}{2}As_4(g) + H_2(g) \rightleftarrows 2GaAs(s) + 2HCl(g)$$

where the letters g, l, and s in brackets identify the gas, liquid, and solid states respectively.

The schematic diagram of a VPE growth apparatus is shown in Fig. 5.19. For $GaAs_{1-x}P_x$ layers to be deposited on the substrate,

arsine AsH_3 and phosphine PH_3 are fed in the mixing chamber together with the *n*-type H_2Se dopant. The mixture is caught up by the stream of carrier gas (hydrogen) and transported to the reaction (center) zone downstream of the gallium zone. The HCl gas flowing over liquid helium transports the volatile gallium monochloride, which is one of the reactants needed for the formation of the $A^{III}B^{V}$ ternary compound. Additions of Zn vapors and H_2 are made into the center zone to form the material of the *p*-type. This VPE technique produces a number of $A^{III}B^{V}$ compounds, including multilayer structures, used for semiconductor emitters: GaAs, GaP, GaN, InAs, InP, AlAs, AlP, $GaAs_{1-x}P_x$, $Ga_{1-x}In_xAs$, $Ga_{1-x}In_xP$, and $Ga_{1-x}Al_xAs$.

Another system of material synthesis by transport reactions uses halides $AsCl_3$ and PCl_3 instead of corresponding hydrogen compounds. The main distinction of this system from that described above is that halide $AsCl_3$ (or PCl_3) is blown into the reactor upstream of the gallium zone because this compound is the source of As_4 (or P_4) and HCl. The chemical reactions proceeding in this system to form GaAs involve the sequential steps of synthesis and transport

$$2AsCl_3(g)+3H_2(g) \rightleftarrows \frac{1}{2}As_4(g)+6HCl(g)$$

$$2HCl(g)+2Ga(l) \rightleftarrows 2GaCl(g)+H_2(g)$$

and epitaxy

$$2GaCl(g)+\frac{1}{2}As_4(g)+H_2(g) \rightleftarrows 2GaAs(s)+2HCl(g)$$

It should be noted that in both systems, the composition of the gas mixture in the center and deposition zones is identical and epitaxial growth is brought about by reducing the substrate temperature below the temperature in the center zone. It is however desirable that the substrate surface should be etched before the epi layer growth. This can be done by raising the temperature in the deposition zone or shifting the equilibrium of the deposition reaction. The HCl excess in the gas stream shifts the epi reaction to the left, causing the gaseous etching. Ease of conversion from etching to epitaxial growth is one of the advantages of the VPE technique.

One more advantage of VPE is that it provides an increased output, 8 to 16 large-size substrates at a time. Fig. 5.20 illustrates the schematic diagram of a VPE growth apparatus for the deposition of $GaAs_{1-x}P_x$ layers on a number of GaAs substrates. The substrates oriented in the ⟨100⟩ direction rest on a holder in the form of a pyramid, which is held at a temperature of 800-825°C during deposition. The boat filled with liquid gallium is kept at 775-800°C with the mixture of HCl and H_2 passed over the boat to form the volatile GaCl reactant and to carry it over to the substrates. The center zone maintained at 900-925°C receives the stream of the carrier gas hydro-

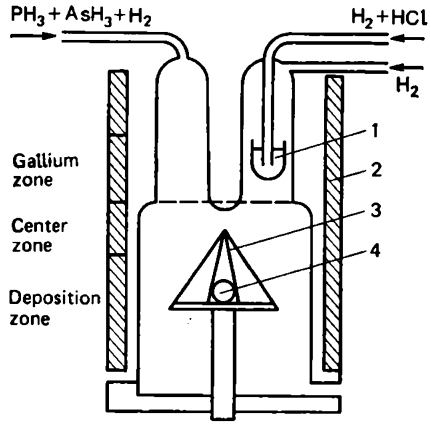


Fig. 5.20. Vapor phase epitaxy growth reactor for deposition of $GaAs_{1-x}P_x$ layers
(1) liquid gallium; *(2)* reactor; *(3)* holder; *(4)* substrates

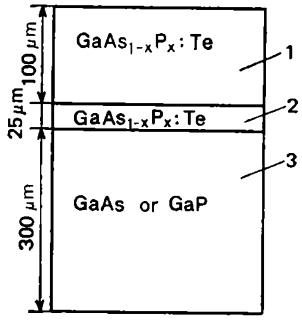


Fig. 5.21. Typical profile of $GaAs_{1-x}P_x$ structure epitaxially grown from the vapor phase
(1) epi layer of constant $GaAs_{1-x}P_x$ composition; *(2)* epi layer of varying composition; *(3)* single crystal substrate

gen delivering AsH_3, PH_3, and the dopant. Fig. 5.21 shows the cross section of a substrate illustrative of the profile of the deposited structure. The epitaxial growth begins with the deposition of a thin GaAs layer on the GaAs substrate at a rate of about 1 μm/min. As the pressure of PH_3 gradually steps up, the deposited film gets more enriched with phosphorus and the rate of growth slowly diminishes. When the phosphorus content reaches the level corresponding to a mole fraction x of 0.4, the growth rate drops to about 0.4 μm/min.

A gradual change in the composition of the epi layer is necessary for two reasons: to obtain a transparent region for light generated by the *pn* junction; to reduce the mechanical stresses and lattice defects, which arise in the course of deposition and cooling of the epi layer since this layer and substrate differ in lattice constant and temperature coefficient of expansion.

Dislocations grow out through the entire deposited layer and spread over largely in the ⟨100⟩ and ⟨211⟩ directions. A high dislocation density reduces the light output of $GaAs_{1-x}P_x$ diodes. In the structure profile, as illustrated in Fig. 5.21, the upper layer has a dislocation density of 10^4-10^5 cm^{-2}, whereas the dislocation density in the middle layer of a varying composition can often be in excess of 10^6 cm^{-2}.

The GaAs and GaP lattice constants at high temperature are not known, but it is found that the lattice distortion at the crystal growth temperature reaches about 4%. Besides, when cooled from

the growth temperature to room temperature, GaAs contracts more than GaP by 15%. Consequently, the GaAs layer deposited under GaP bends the latter so that it becomes convex toward the viewer.

Bending stresses in the layers of Fig. 5.21 strongly depend on the thickness of both the transition layer and the grown layer. Where the composition of the layer smoothly varies 20 to 40 μm in depth, the structure may have fairly low bending stresses if, however, the next layer 40 to 80 μm thick is homogeneous in composition. For the given thickness of the substrate and the transition layer there is an optimum thickness of the homogeneous layer at which the bending stress is at a minimum.

One of the main advantages of VPE is that it can produce ternary systems with a smooth variation of the composition. For preparation of GaP as well as GaAs, this technique provides two types of volatile gallium compounds—gallium monoxide and galides. Hydrogen with water vapors serves to form and transport Ga_2O. In contrast to LPE, VPE can introduce far larger concentrations of nitrogen to enhance the recombination, but in principle proves ineffective and involves difficulties as regards the introduction of oxygen. For this reason, the VPE process can be attractive for the fabrication of efficient green GaP LEDs but may not be suited to prepare the materials for red GaP LEDs.

In the text above, we enumerated the benefits accrued in the application of VPE. Let us now point to its limitations.

1. The crystal defect density in clearly defined layers of binary compounds is about the same as that in the substrate, but is higher in solid solution layers of a varying composition.

2. The quantum efficiency of LEDs with the active layer at the substrate-epi layer interface is commonly low, therefore the *pn* junctions formed in this configuration are unsuitable for LEDs.

3. The deposition of epi layers necessitates the use of only those chemical elements which form volatile compounds.

Consequently, this method is disadvantageous for the preparation of aluminum compounds. From the above considerations it follows that VPE is a convenient process to produce certain LEDs, in particular $GaAs_{1-x}P_x$ diodes. But red LEDs using GaP and $Ga_{1-x}Al_xAs$ should preferably be made by the LPE technique.

5.2.4. PN junction fabrication by liquid phase epitaxy (LPE). This technique can give semiconductor emitters of the highest quantum efficiency and finds extensive uses. Liquid phase epitaxy is a specific case of crystal growth from the solution. The basic steps of the LPE process are the following. First, a liquid metal, for instance Ga, is saturated at an elevated temperature with a semiconductor compound such as GaAs. The GaAs single-crystal substrate is heated to the same temperature as the melt when the dissolved material is not yet uniformly distributed in the melt (Fig. 5.22*a*). The substrate

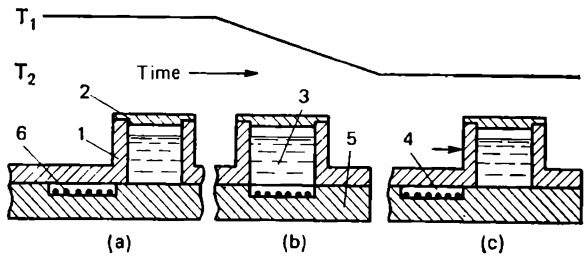


Fig. 5.22. Liquid phase epitaxy growth stages
(*a*) saturated melt and substrate heated to initial temperature T_1; (*b*) substrate moved into contact with melt, melt temperature lowered to initiate crystal growth from oversaturated melt; (*c*) substrate removed from melt; (*1*) boat; (*2*) cap; (*3*) Ga-GaP-dopant melt; (*4*) epi layer; (*5*) base; (*6*) substrate

and the melt are then brought in contact (Fig. 5.22*b*). If the substrate and melt differ in temperature, a fraction of the substrate dissolves in the melt or a fraction of the melt precipitates as a solid phase on the substrate. Upon cooling, the epitaxial GaAs layer grows on the substrate. Cooling causes oversaturation of the melt with As atoms, so arsenic precipitates in the solid phase to form GaAs. Ideally, the epitaxial growth proceeds only on the substrate to yield the layer by epitaxy from the liquid phase. The crystal growth can be arrested by mechanically moving the substrate away from the melt (Fig. 5.22*c*).

Comparing this process with VPE described above allows us to conclude the following. The crystal grows at the solid-liquid interface, so it has a lower dislocation density than the substrate. This is one of the most important merits of the LPE process. Besides, LPE offers optimum conditions for decreasing the gallium vacancy density in the grown crystal. That is why the lifetime of minority carriers in LPE-grown crystals is usually about 500 times higher than that in crystals prepared by other methods. A decreased lifetime commonly entails a reduction of the internal quantum efficiency by a factor of over 100. Theory confirms that the epi layers grow in direct proportion to the diffusion constant of the dissolved material and the time of growth, but inversely with the square of the melt thickness. So, for the deposition to be more effective, it is wise to reduce the thickness of the melt layer. The LPE process requires a large number of source materials, i.e. it necessitates the preparation of saturated melts for each cycle of growth. By contrast, in the VPE process, many layers can be grown from one source. The productive capacity of an LPE system is generally much lower than that of a VPE system. Many shortcomings of the liquid phase epitaxy technique can be overcome by using thin melt layers, less than 1 mm thick.

LPE arrangements. There are three basic types of apparatus and thus three variations of the LPE technique to prepare the desired

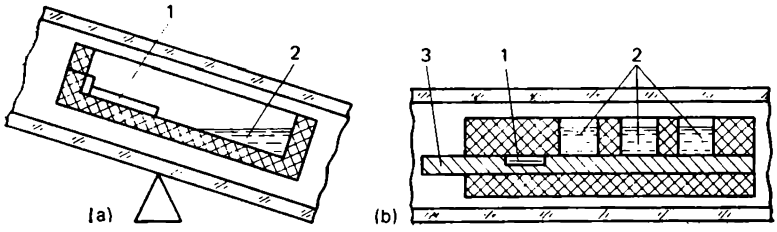


Fig. 5.23. Systems with tipping boat (*a*) and multiple-bin boat (*b*)
(*1*) substrate; (*2*) melt; (*3*) movable insert

materials: a tilting-boat apparatus where the melt comes in contact with the substrate by tipping the boat (tilting-boat technique); vertical-boat apparatus where the substrate is dipped into the melt contained in a vertical boat (dipping technique) and multiple-bin apparatus where several epi layers are grown sequentially as the substrate is brought in contact with different solutions (multiple-bin technique).

Consider the principle of operation of the first system. The layer here is grown from a Ga melt to which is added GaAs, GaP, or their mixture, and a dopant such as Zn. The melt is positioned at one end and the substrate at the other end of the boat (reactor) tilted so that the melt makes no contact with the substrate (Fig. 5.23*a*). The boat is set up in a uniformly heated zone of the furnace's quartz tube through which hydrogen is forced to flow. As the temperature reaches the specified value, the boat is tipped so that the melt covers the substrate. The heater power is then stepped down to cool the boat at the required rate. The epi layer grows during slow cooling. The melt composition is so chosen that it comes to, or close to, equilibrium before the melt is brought in contact with the substrate. If the melt is near the equilibrium state, the substrate partially dissolves and its surface becomes free of imperfections. Recrystallization then proceeds on the highly polished substrate during subsequent cooling. The epi layer is grown to the desired thickness by controlling the range of growth temperatures and the rate of cooling. The boat is held in this temperature range for a definite time and then returned to its initial position to allow the melt to flow off the substrate.

This technique can produce two layers, the *n* layer using the Te dopant and then the *p* layer by adding the Zn dopant. Two approaches apply for the purpose—overcompensation in the melt and doping from the vapor phase. By the first approach, when the melt still covers the substrate and the crystal growth process continues as the *n* layer becomes 15 to 20 μm thick, an acceptor is added to the melt to overcompensate it. This approach allows a continuous growth

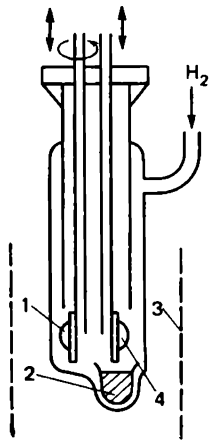


Fig. 5.24. Liquid phase epitaxy growth apparatus with vertical immersion of substrates (*1*) GaAs substrate; (*2*) melt; (*3*) furnace; (*4*) GaAs source

of the *p* layer on the *n* layer. The second approach calls for a sufficiently thin melt layer to cover the substrate. Here too doping is done during the uninterrupted layer deposition by introducing H_2S or Zn vapors into the reactor. The advantage of the first technique is that the *pn* junction boundary has no contact with the atmosphere and requires no intermediate polishing of the substrate surface.

The vertical LPE growth apparatus for the deposition of layers by immersing substrates into the melt is shown in Fig. 5.24. The substrates suspended on the holder over the graphite or corundum melt-filled boat are dipped into the melt at a definite temperature and for a specified time to allow the epi layers to grow in the same manner as do the layers in the first process described above. As the specified time of growth elapses, the substrates are withdrawn from the melt. The dipping technique makes it possible to dope the melt in the course of growing with donor and acceptor impurities and to easily control the thickness of epi layers. The technique is also convenient in the case when an oxide film formed on the melt surface inhibits contact with the melt.

The multiple-bin apparatus applies for the sequential deposition of several epi layers of different compositions, if need be, in the single cycle of growing. Fig. 5.23*b* illustrates one of the design versions of the graphite multiple-bin boat of the pencil-box type. The movable graphite insert with a hollow to accommodate the substrate serves as a bottom of the boat. The boat with bins containing gallium and requisite III and V materials is put into the furnace's quartz tube. When the melts come to equilibrium on heating, the insert with the substrate is moved sequentially to form *n* and *p* layers and change the composition of growing epi layers should the need arise. This technique is particularly suitable for growing $Ga_{1-x}Al_xAs$ layers, which are otherwise difficult to produce because of a high diffusivity of aluminum. In addition to growing multilayer structures, this technique can produce layers on several substrates at a time, thus increasing the output.

5.2.5. Ohmic junction fabrication. The ohmic junction is a metal-semiconductor contact whose resistance is independent of the direction of current in the entire range of current values. An ideal ohmic

junction is practically impossible to obtain. An ohmic junction is made by forming a metal contact to a strongly doped semiconductor. As the majority carrier concentration grows, the width of the depletion layer shortens, so the carriers begin to tunnel through the layer. The general approach to the fabrication of ohmic junctions between metals and all $A^{III}B^{V}$ semiconductor materials is to deposit a metal on a heavily doped semiconductor region to enable carrier tunneling.

A high carrier concentration in the semiconductor can be attained by, for example, diffusing Zn in the same manner as is done in the fabrication of *pn* junctions. The surface concentration of diffused Zn is usually so high that any metal, say, gold or aluminum, will do well to accomplish the ohmic junction. A low level of semiconductor doping can be raised in the process of formation of the ohmic junction. For this the alloy used should contain the required dopant—the alloying element dissolved in its base metal. A number of methods are available for alloy deposition: thermal (vacuum) evaporation, ion-beam evaporation, etc. The deposited alloy should be burnt in to introduce dopants into the semiconductor. The following requirements must be met to produce adequate junctions: (1) the selected metal must readily wet the semiconductor; (2) the materials used must ensure the lowest possible stresses at the interface; (3) a transparent semiconductor, for instance GaP, must weakly absorb light because a large share of light undergoes multiple internal reflection before it escapes the semiconductor, and (4) the fabrication process must be consistent with the subsequent operations of mounting, wire bonding, etc.

These requirements can only be filled using a binary or ternary alloy rather than a single metal. The base metal (solvent) must have a low electrical resistance and adequate mechanical strength to provide a reliable joint in bonding of connection wires. The second metal is an alloying element (solute), and the third metal serves to improve, for example, wettability and other properties. The donors are Si, Ge, Sn, and Te for GaAs and Si, Sn, and Te for GaP. The acceptors for GaAs and GaP are Be, Mg, and Zn. The principal metal is commonly gold. The Au-Ge and Au-Zn alloys (5 to 25% Ge and 4 to 10% Zn) are used for *n* regions, and the Au-Zn alloy with 1 to 25% Zn for *p* regions. Au-Ge and Au-Si alloys form ohmic contacts to GaAs, GaAsP, GaInP, and GaP. Ohmic contacts are formed by deposition in chambers pumped to a vacuum of 1.3×10^{-3} Pa at a substrate temperature of 200-300°C. Nickel is sometimes grown over the alloy to afford ease of wire lead bonding. The junction formed just after evaporation of the alloy on to the substrate shows a nonlinear *VI* characteristic and needs firing at the optimum temperature for the given compound. The optimum burn-in (firing) temperatures for GaAs, GaP, and GaAsP are 450°C, 590-630°C, and 530-580°C respectively. Firing at higher temperatures causes an increase in the

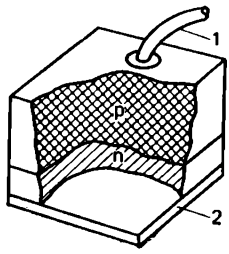


Fig. 5.25. GaP LED structure (*1*) fused-in Au-Zn contact; (*2*) deposited Au-Si contact land

junction resistance. A change in the donor concentration from 2×10^{17} to 10^{18} cm^{-3} changes the junction resistivity from 10^{-3} to 10^{-4} Ω cm^{-2}, which is the highest value for GaAlAs and the lowest for GaAsP. With a change in the donor concentration N_d, the resistivity of these materials varies linearly as $\sqrt{N_d}$.

Au-Zn alloys form a good ohmic contact without firing only to GaAsP. In GaAlAs the junction resistivity decreases with growing burn-in temperature and reaches 10^{-5} Ω cm^{-2} at 450-500°C. GaP shows a minimum resistivity of 6.5×10^{-4} Ω cm^{-2} if the junction is burnt in at 600°C. The junction resistivity on the *p* region side is a factor of 10 to 100 lower than that on the *n* region side.

Silver-base alloys find rarer uses. Ag-Sn alloys form ohmic contacts to *n*-GaAs at a burn-in temperature of 450°C. The Ag-Ge-In and Ag-Ge-Ni alloys burnt in at 850°C for 1 min give good ohmic contacts. The Ag-Te or In-Ni alloy is used for *n*-GaP, and Zn-Ni alloy for *p*-GaP.

Aluminum is a widely used metal to produce aluminum contacts 1 or 2 μm thick to the *p* regions of solid solutions GaAsP and GaAlAs. An aluminum junction requires firing at 550°C. As the firing temperature grows the junction resistivity drops off; it is equal to $2 \times \times 10^{-5}$ Ω cm^{-2} and 1.5×10^{-6} Ω cm^{-2} at 600 and 500°C, respectively, for CaAlAs and GaAsP. Aluminum does not interface well with GaP.

Light absorption by ohmic contacts. It should be kept in mind that junctions tend to heavily absorb light after firing. A change in the absorption capacity stems from the intermetallic compounds formed in the structures, the effective bandgap energy of which is lower than that of semiconductors.

Look at a simple structure of the GaP LED where the semiconductor has a nearly cubic form (Fig. 5.25). One side of the cube is entirely covered with metal and the opposite side has a fused-in ohmic contact. The semiconductor surface becomes relatively rough after the manufacturing operations. Photon emission in the *pn* junction region is isotropic in character. Almost 50% of generated photons move in the direction of the metallized surface. On the average, each photon undergoes multiple internal reflection before it leaves the diode structure. Most photons bounce back from the metallized surface to be subsequently absorbed in the structure. The metal can thus cause a substantial loss of light and a decrease in the output efficiency. The newly-formed gold layers reflect well the red and

green light generated by a GaP LED, but strongly absorb the light after firing. So, the process employed to reduce the junction resistivity and provide good mechanical strength is conducive to a lower external quantum efficiency of LEDs. The tradeoff between the opposing requirements largely calls for the deposition of a dielectric film, for instance SiO_2, on the GaP-metal interface.

The dielectric film on the metal-semiconductor interface has a dual purpose. First, it changes an internal reflection angle on the GaP-metal interface. A smaller TIR angle causes a larger share of light propagating in arbitrary directions to experience TIR. Typical values of the refractive index for SiO_2 and GaP are 1.5 and 3.4 respectively, so the TIR angle is equal to arcsin $1.5/3.4 = 23.7°$. Second, the dielectric film retains the reflection capacity of the deposited metal so that the share of light emanating within the TIR angle reflects back into the diode. The surface dielectric film about 10 nm thick is sufficient to retain the reflectance. To achieve the aim, the semiconductor surface can initially be treated in an etchant such as boiling hydrogen peroxide. Gold- and silver-base alloys are found to adhere well to pure dielectric layers.

5.2.6. Typical processes of LED fabrication. The main parameters of source materials for LEDs are the degree of crystallinity, conductivity type, resistivity, carrier concentration, carrier mobility, and absorptance. These parameters are measured by the known methods of semiconductor engineering. The materials are tested for the stability of parameters and duly selected for the fabrication of LEDs. At the beginning of this chapter we briefly looked into the basic processes harnessed for the manufacture of optoelectronic devices. Basic steps in producing LEDs include: (1) wafer preparation from the starting material; (2) *pn* junction formation; (3) ohmic junction formation; (4) scribing to separate wafers into chips; (5) mounting, bonding, and packaging.

The preparatory stage includes the mechanical and chemical treatment of wafers, the aim being to reduce the wafer thickness to 0.2 or 0.3 mm and give the desired polish. From the viewpoint of material economy and absorption reduction, the wafers should be as thin as possible. However, the wafers thinner than 0.2-0.3 mm lose mechanical strength and readily crack up.

The next stage—*pn* junction fabrication—involves diffusion, vapor-phase epitaxy (VPE), and liquid phase epitaxy (LPE)—the processes described in detail above. Here we only touch in passing on these techniques involved in LED fabrication.

The VPE process affords a high degree of reproducibility, readily lends itself to control, and, besides, can treat thousands of LED structures simultaneously. The equipment required for the purpose is the same as that used for production of silicon ICs.

The LPE process is more laborious than VPE, calls for more costly

super-pure materials and more complex equipment. On the other hand, the LPE technique is well adapted to processing a much larger number of LED materials. For many $A^{III}B^{V}$ compounds it is the only technique that can produce good emitting structures.

The process of diffusion of acceptor and donor impurities from the vapor phase secures an excess vapor pressure, thus preventing semiconductor compounds from decomposition. The technique is adequately fit for the subsequent photomasking processes. However, the diffusion process is poorly controllable and cannot provide LED structures with a good set of parameters.

The basic materials used for modern semiconductor photoemitters and types of technique for their manufacture are listed in Table 5.1,

Table 5.1. **Characteristics of Compounds for Semiconductor Photoemitters**

Substrate	Material	Technique		Wavelength, λ	Color	Quantum efficiency, %
	pn junction	*n* region	*p* region			
GaAs	GaAs : Si, Zn	Substrate	Zn diffusion	appr. 940	IR	3.5
	GaAs : Si	LPE	LPE	appr. 970	IR	32
	$Ga_{1-x}Al_xAs$	LPE	Zn diffusion	> 700	IR	13.5
	$GaAs_{0.7}P_{0.3}$	VPE	Zn diffusion	—	Red	0.7
	$GaAs_{0.6}P_{0.4}$	VPE	Zn diffusion	appr. 680	Red	0.2
GaP	$GaAs_{0.6}P_{0.4}$	VPE	Zn diffusion	appr. 680	Red	0.09
	$GaAs_{0.35}P_{0.65}$:N	VPE	Zn diffusion	appr. 630	Red-orange	0.65
	$GaAs_{0.1}P_{0.9}$: N	VPE	Zn diffusion	appr. 580	Yellow	0.08
	GaP : Zn, O	LPE	Zn diffusion	637	Red	1.5
	GaP : Zn, O	LPE	LPE	637	Red	12
	GaP : N	LPE	Zn diffusion	—	Yellow	0.1
	GaP : N	VPE	Zn diffusion	574	Yellow-green	0.1
	GaP : N	LPE	Zn diffusion	570	Yellow-green	0.1
	GaP : N	LPE	LPE	568	Yellow-green	0.33
	GaP : N	LPE	LPE	570	Yellow-green	0.7

which also presents the color of luminescence, peak emission wavelength λ_{max}, and external quantum efficiency η. From the comparison of the data it is seen that the emission power (quantum efficiency) generally grows with wavelength λ_{max}. Therefore, the emitters designed for application in the invisible IR region largely use *n*-GaAs and GaAlAs with λ_{max} of 0.80 to 0.95 μm. For optical communication lines, the radiation at a wavelength of 1.2 to 1.3 μm provides for optimum transfer of signals. The LEDs designed for these lines are made from GaSb and ternary compounds based on GaSb.

An ohmic contact to the semiconductor substrate is made by depositing a 1-μm film of gold, nickel, or silver of a high work function.

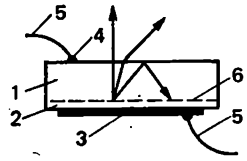


Fig. 5.26. Structure of planar photoemitter
(*1*) *n* region; (*2*) *p* region; (*3*) ohmic contact to *p* region; (*4*) ohmic contact to *n* region; (*5*) leads; (*6*) *pn* junction (arrows show paths of rays)

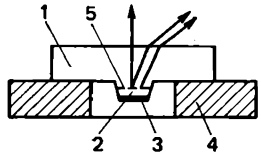


Fig. 5.27. Photoemitter mesa
(*1*) *n* region; (*2*) *p* region; (*3*) ohmic contact to *p* region; (*4*) ring electrode to provide contact to LED base; (*5*) *pn* junction

To the base metal is added a donor impurity if the aim is to form a low-resistivity contact to the *n*-type material, or else an acceptor is added if the contact to the *p*-type material is required. In some cases ternary alloys are used to form contacts of improved mechanical properties. An alloy film is grown by vacuum evaporation or chemical deposition, and the structure so formed is heat treated. One of the contacts is sometimes made transparent to reduce light absorption. For this a film of tin dioxide or SiO_2 is deposited on the semiconductor surface. In other cases a contact on the base region is made in the form of a bump rather than as a bonding pad.

After forming contacts, the slice (wafer) is cut into chips 0.5 $\times$ 0.5 to 2 $\times$ 2 mm in size, each incorporating one LED. The chip is mounted on the header with its bonding pad resting on the conductor pattern of the header. In the next operation, a strong bond is made between chip and header, for example, by thermocompression. The other contact on the chip is connected to the header terminal with a jumper—a fine metal wire. The mounted chip is now ready for encapsulation in a metal envelope. Plastic encapsulation is now very popular. Here the chip with wire leads bonded to its junctions is coated with a liquid transparent plastic, which polymerizes into a ball encapsulating the chip.

Figure 5.26 illustrates the structure of a LED of the simplest design. The light generated by the *pn* junction can go out through the upper face into the atmosphere only if $\sin \varphi < 1/n$, where n is the refractive index of the material and φ is the angle of incidence. Otherwise, the rays would be subjected to TIR and absorbed in the base material or the lower ohmic junction. This explains why the throughput efficiency of such emitters does not exceed 2-4% even if the internal quantum efficiency is close to unity.

The ratio of the external to the internal quantum efficiency can be raised using mesa-structures, or mesas. One stage in the fabrication of such a structure for a mesa diode is to coat the *p* layer with wax, if the LED base is made from an *n*-type material, and etch the chip

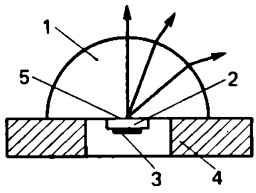


Fig. 5.28. Hemispherical structure of photoemitter (*1*) *n* region; (*2*) *p* region; (*3*) ohmic contact to *p* region; (*4*) ring electrode to form contact with LED base; (*5*) *pn* junction

to lower the junction (Fig. 5.27). The light emitted by the mesa diode then reflects from the side faces and falls on the plane surface of the base almost at a right angle. The light output thus increases 2- or 3-fold. This technique of raising the external quantum efficiency is rather simple from the technology viewpoint, but cannot completely exclude the loss due to TIR.

Shaping the base as a hemisphere with the *pn* junction in its central area increases the external quantum efficiency almost ten-fold as against the efficiency of a flat diode (Fig. 5.28).

Solid solutions go for the batch manufacture of diode matrixes. The first fabrication step involves the VPE growth of a layer with a smoothly varying composition along the x coordinate on the *n*-type GaAs substrate. A smooth variation in the composition reduces the number of dislocations that result from the difference in crystal lattice constants between the substrate and growing layer. To achieve the aim, it is necessary to introduce thoroughly proportioned vapors of As, P, Ga, and Se into the reaction chamber.

After forming the layer with a varying composition, a GaAsP layer is grown and doped with selenium to a donor concentration of 2×10^{18} cm^{-3}. Next zinc chloride is added in an amount enough to reach an acceptor concentration of 4×10^{18} cm^{-3} in the growing layer. This layer should be as thin as possible, so that it can absorb a lower amount of light generated by the *pn* junction. In the subsequent step, the supply of phosphorus vapors is stepped up with the amount of arsenic vapors reduced to form the base region that can weakly absorb light. In the final step, the zinc chloride content is raised to produce an enhanced-conductivity layer which enables the current to uniformly spread over the junction surface.

After formation of the mesas, a number of cylindrical protrusions are left on the substrate, each of which is an emitting element. The substrate is then covered with a thin SiO_2 layer which serves as a passivation coat.

The matrix so produced is molded over into liquid glass under pressure. Next holes for contacts are etched in the glass opposite the cylindrical protrusions, the starting substrate is removed, and contacts with interconnections are provided in the process of metallization.

The main structures of semiconductor photoemitters are shown in Fig. 5.29. The epitaxial GaAs(Si) structure, as illustrated in Fig. 5.29*a*, provides the highest quantum efficiency due to an impro-

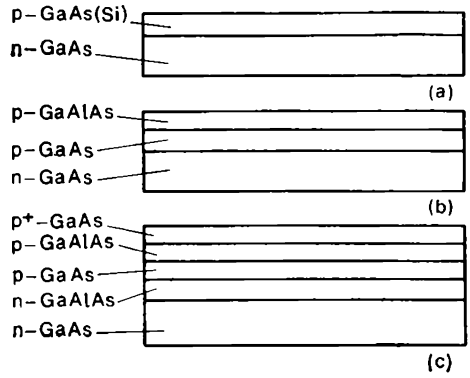


Fig. 5.29. Typical photoemitter structures
(*a*) epitaxial GaAs structure; (*b*) one-sided heterostructure; (*c*) two-sided heterostructure

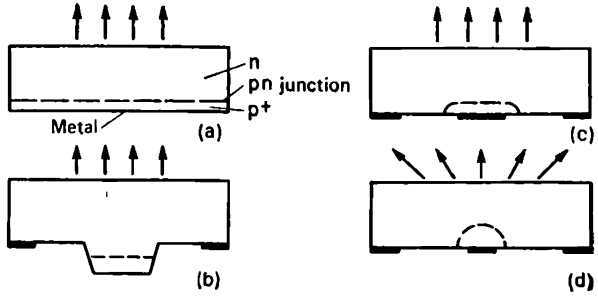


Fig. 5.30. Arrangement of the active region on the LED substrate
(*a*) planar structure; (*b*) mesa; (*c*) planar structure; (*d*) hemispherical structure

ved extraction of light with a minimum of loss. The region of photon generation is rather large, for which reason the switching time t_{sw} is long, 10^{-6}-10^{-7} s. The one-sided two-layer structure of Fig. 5.29*b* offers all the advantages of heterostructure devices.

In the two-sided three-layer structure of Fig. 5.29*c*, the internal quantum efficiency reaches a theoretical limit owing to elimination of the effects of absorption and spreading of carriers from the active region.

The emitting region of a LED can be sited on the chip in different ways (Fig. 5.30). One commonly strives to minimize its area with the aim to attain the desired values of emission parameters at lower values of the forward current. To reduce the loss of light in the course of its extraction from the chip, i.e. to increase the optical efficiency of a LED, the chip is sometimes shaped to a more complex form different from the parallelepiped configuration. A semiconductor is optically denser than the air. The refractive index of, say, gallium

arsenide is equal to 3.3. For this reason a major portion of light reflects from the air-semiconductor interface into the chip and is finally lost through absorption. Shaping the semiconductor of the LED as a sphere increases the optical efficiency by a factor of 15-25.

5.3. Fabrication of Photodetectors

5.3.1. Materials. Semiconductors are the main materials used for the fabrication of solid-state photodetectors (PDs). Photosensitive semiconductors can be in single-crystal, polycrystalline, and amorphous states. Table 5.2 presents the basic parameters of semiconductor materials most widely used for the fabrication of PDs.

Silicon, which is prepared and processed by rather well developed technology, is the basic material for *pn* junction photodetectors.

Table 5.2. **Photodetector Materials**

PD type	Material	λ_{max}, μm	t_{sw}, s
pn junction	Si	0.8-0.9	10^{-10}
	GaAs	0.80	10^{-10}
	Ge	1.55	10^{-9}
	InAs	3.30	10^{-10}
photoresistor	CdS	0.57	10^{-2}
	CdSe	0.67	10^{-2}
	PbS	2.50	10^{-4}

As mentioned earlier, the photon-coupled pair (PCP) composed of a silicon *pin* photodiode and a GaAs LED enjoys extensive uses today; the pair of these devices is well compatible in both spectral and transfer characteristics.

GaAs and GaAlAs photodetectors are the best match for most effective LEDs such as GaAlAs heterostructure devices. These photodetectors are also of the heterostructure type.

The use of silicon PDs for operation at the emission wavelength of solid-state lasers is little effective. The YAG laser, for instance, emits at 1.06 μm. Solid-state lasers mainly operate with PDs based on Ge, InAs, and ternary compounds InGaAs and GaAsSb.

The main materials for photoresistors are CdS, CdSe, and mixtures of these binary systems. Copper atoms introduced in these semiconductors serve as response centers: they sharply increase the lifetime of electrons in the conduction band.

5.3.2. Fabrication of pin photodiodes. Consider the basic techniques of manufacture of photodetectors. It should be pointed out that diode structures form the basis of most varieties of photodetectors employed in modern optoelectronic devices. Among photodiodes

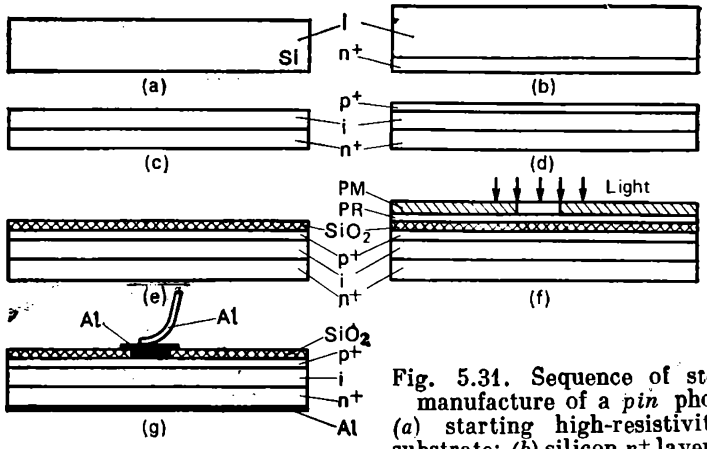


Fig. 5.31. Sequence of steps in the manufacture of a *pin* photodiode (*a*) starting high-resistivity silicon substrate; (*b*) silicon n^+ layer epitaxial growth; (*c*) *i* layer thinning; (*d*) silicon p^+ layer epitaxial growth; (*e*) silicon oxidation; (*f*) photomasking; (*g*) metallization and lead bonding

used for optoelectronic devices, the basic and most promising designs include a silicon *pin* diode, Schottky-barrier (metal-semiconductor) diode, and heterode (heterodiode).

A typical process of manufacture of *pin* photodiodes comprises a few basic steps (Fig. 5.31). First, a low-resistivity 0.02-Ω m n^+ silicon layer 30 to 50 μm thick is epitaxially grown from the vapor phase on the high-resistivity slice with ρ ranging from 0.02 to 0.10 kΩ m (Fig. 5.31*a*, *b*). Next the slice is inverted to lap the high-resistivity *i* layer of the starting silicon material to a thickness of about 50 μm (Fig. 5.31*c*). A low-resistivity epitaxial p^+ layer 2 to 4 μm thick is then grown on top of the *i* layer and coated with SiO_2 (Fig. 5.31*d*, *e*). Following this, the slice undergoes a masking step to selectively etch away SiO_2 and open windows for contacts to the p^+ layer. The final sequence of steps in the manufacture of *pin* photodiodes is as follows: deposit a thin aluminum (or gold) layer over the entire slice; etch away the unwanted portions of the aluminum layer by photoengraving to leave behind the desired interconnection pattern; scribe the slice to separate it out into chips; protect side surfaces of each chip; mount each chip on the header; interconnect with the lead frame of the header; encapsulate in a metal envelope.

Ion implantation or diffusion into windows opened in the SiO_2 film is another possible approach to form a low-resistivity p^+ layer. The ratio between the dimensions of the structure layers is such that the active *i*-type region accounts for 90% of the energy absorbed in the structure; the share of light energy absorbed in the p^+ and n^+ regions does not exceed 5-10%. An antireflection SiO_2 coat formed on the photodiode surface reduces the reflection loss and increases the efficiency of photoelectric conversion.

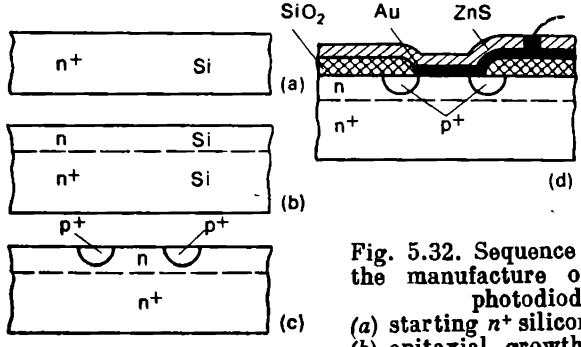


Fig. 5.32. Sequence of steps in the manufacture of Schottky photodiode
(*a*) starting n^+ silicon substrate; (*b*) epitaxial growth of silicon *n* layer; (*c*) local diffusion to form p^+ guard ring; (*d*) silicon metallization, ZnS antireflection coat deposition, and lead bonding

From the viewpoint of economy and versatility, the *pin* photodiode fabrication technology has several principal drawbacks. The starting high-resistivity silicon calls for high purity, which increases the product cost. Also, a high purity of the basic region is responsible for a low height of the potential barrier. That is why the photo-emf of these devices cannot be higher than 0.35-0.45 V. A low potential barrier is also the cause of increased leakage currents and lower radiation stability. High-resistivity silicon is less homogeneous and more subject to uncontrollable changes, so the reproducibility of *pin* structure parameters is poorer than is the case for *pn* structures. The matter of particular concern is that *pin* structures are technologically incompatible with integrated circuits made up of thin doped layers. This circumstance complicates the accommodation of a photodiode and hf transistor on the same chip.

5.3.3. Fabrication of Schottky-barrier photodiodes. Fig. 5.32 illustrates the basic sequence of steps in the manufacture of Schottky photodiodes. The first manufacturing step begins with the VPE growth of a thin high-resistivity silicon layer on the starting n^+ silicon slice. The planar technique then produces a *p*-type ring in this layer (guard ring). After the surface of silicon inside the guard ring is thoroughly cleaned, a semitranslucent gold film about 0.01 μm thick is deposited on top of the silicon layer and is then given an antireflection zinc sulfide coat. The guard ring serves to reduce the leakage current and raise the breakdown voltage.

The Schottky barrier fabrication technique offers several advantages over other techniques harnessed to form various photodiode structures. This technique relies on the standard planar technology, which produces metal-semiconductor contacts at temperatures that are relatively low from the viewpoint of semiconductor fabrication practice. Besides, Schottky photodiodes easily interface with optical ICs.

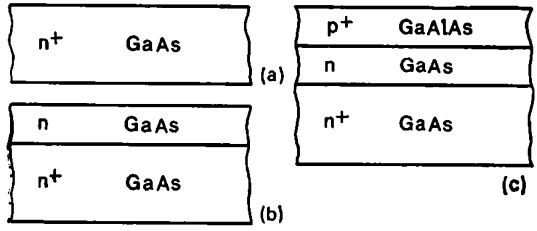


Fig. 5.33. Process of manufacture of the photodiode structure with a heterojunction
(*a*) starting n^+-GaAs substrate; (*b*) epitaxial growth of GaAs *n* layer; (*c*) epitaxial growth of p^+-GaAlAs solid solution

5.3.4. Fabrication of semiconductor heterodiodes. To get an idea of the manufacturing process, look at the basic steps of the formation of a GaAs-GaAlAs structure (Fig. 5.33). The starting material of the substrate is n^+ GaAs with a donor concentration of 10^{18} cm^{-3}. First, an undoped *n*-type GaAs layer is epitaxially grown from the liquid phase on the substrate. The donor concentration of this layer ranges from 10^{15} to 5×10^{15} cm^{-3}. The next step of LPE growth provides a p^+ GaAlAs layer with an acceptor concentration of about 10^{18} cm^{-3}. This upper layer transmits light practically without any loss, and the middle *n*-type GaAs layer absorbs this light which triggers the generation of free photocarriers. The holes generated in *n*-type GaAs move into the *p* region. The thickness of the *n* region must be chosen such that it can absorb all of the incident light. At the 0.85 μm wavelength of incident radiation, an *n*-type region 20 μm thick does well for the purpose. A high purity of the *n* region ensures a reduced loss due to photocarrier recombination. The available techniques can now produce high-quality heterojunctions with a low density of surface states at the layer boundaries. This feature also aids in reducing the parasitic recombination loss.

It is of importance to stress the point that the heterode structure allows for adjusting the spectral capability of the photodetector in a wide range. By selecting the adequate pair of semiconductors, it is possible to reach a high efficiency of the PD in any spectral region of interest. A wide range of available materials offer considerable scope for optimum design of heterojunction PDs harnessed to pick up signals at different wavelengths, in contrast to silicon photodiodes where χ_0-λ relation limits their operating frequency range.

Besides, since there is a free choice of the requisite materials for the base, it is possible: (1) to produce selective photodiodes where the *n* region (base) and transparent p^+ region (window) have close values of the bandgap energies; (2) to increase the values of photo-emf, which, for example, reach 0.8 to 1.1 V for GaAs-GaAlAs heterodes and are twice or thrice the values of silicon diodes; and (3) to

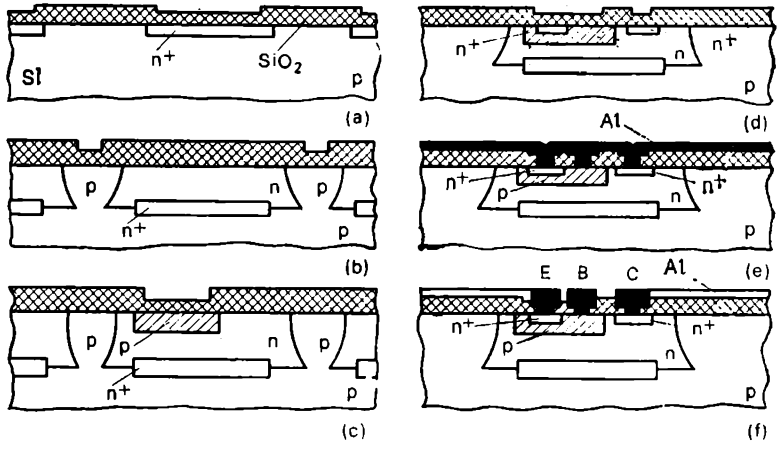


Fig. 5.34. Sequence of steps in the manufacture of a planar phototransistor (*a*) diffusion to form buried n^+ layer; (*b*) isolation diffusion; (*c*) base diffusion; (*d*) emitter diffusion; (*e*) metallization; (*f*) photomasking to form aluminum interconnection pattern

reduce reverse currents and extend the range of operating temperatures. Another factor in favor of heterostructures from the technology viewpoint is that they are compatible with dielectric waveguides of optical ICs.

The main limitations to heterodes are a complex manufacturing process and difficulties involved in the selection of the adequate pair of semiconductor materials. Such a pair must be a perfect match in lattice constants, TCs of linear expansion, etc.

5.3.5. Fabrication of phototransistors. The basic type of photodetector that provides for current amplification is a phototransistor fabricated by the standard planar technology. The main feature specific to a planar phototransistor is that all its working layers come out on one side of the substrate with all the electrodes arranged in the same plane. Figure 5.34 illustrates the sequence of steps in the manufacture of a planar *npn* transistor with a buried layer. The first manufacturing operation involves thermal oxidation of the starting *p*-type silicon wafer to cover it with a SiO_2 film which performs the functions of a passivation layer and a mask to define windows for successive diffusions. The further sequence of operations is as follows:

(1) first photomasking to open windows in the oxide for an n^+ layer;

(2) first diffusion to form the n^+ layer using arsenic or antimony as a dopant (Fig. 5.34*a*);

(3) oxide etching from the entire surface of the substrate;

(4) epitaxial n layer growth;
(5) oxidation;
(6) second photomasking to open windows in the oxide for the second (isolation) diffusion;
(7) second diffusion to form isolating p layers and n-type islands using boron as a diffusant (Fig. 5.34*b*);
(8) third photomasking to open windows in the oxide film for the third (base) diffusion;
(9) third diffusion to form p-base regions by driving in boron (Fig. 5.34*c*);
(10) fourth photomasking to open windows for the fourth (emitter) diffusion;
(11) fourth diffusion to produce n^+ emitter region by driving in phosphorus (Fig. 5.34*d*);
(12) fifth photomasking to open windows for ohmic contacts;
(13) aluminum deposition over the entire substrate surface (Fig. 5.34*e*);
(14) sixth photomasking to open windows in the photoresist mask and thus define the interconnection pattern;
(15) aluminum etching to remove the unwanted portions, leaving behind the interconnection pattern (Fig. 5.34*f*);
(16) thermal treatment to fuse aluminum into silicon;
(17) assembling and packaging.

In addition to the above manufacturing steps, the entire fabrication process also includes many operations called upon to clean and wash the substrates and also to deposit and remove the photoresist.

Consider some manufacturing steps in more detail. It is in principle possible to use a p-type substrate with an epi-n layer deposited over its surface rather than a monolithic n-type slice as the starting material. The second diffusion (7th step) called isolation diffusion, which forms collector layers, will then be made through the entire epi layer 10 to 15 μm thick as far as the p substrate. The diffusion takes 4 to 6 hours to drive the dopant to such a depth.

The isolation diffusion thus produces n-type islands in the pockets separated from each other by pn junctions formed in the epi layer during this diffusion. But the transistors formed in these pockets have a very high lateral resistance r_C, 100 ohms and above. It is impracticable to decrease the resistivity of the epi layer since the breakdown voltage of the collector junction then drops off and the collector capacitance grows. A practical way of reducing the problem is to form a buried n^+ layer at the bottom of the pocket (Fig. 5.35). The collector current then passes through the low-resistivity n^+ layer and the resistance r_C becomes much lower. The buried layer is formed by diffusion before deposition of the epi layer (2nd step). During the subsequent epitaxy (4th step) the donor atoms excited by heat diffuse into the growing epi-n layer. To guard against an exces-

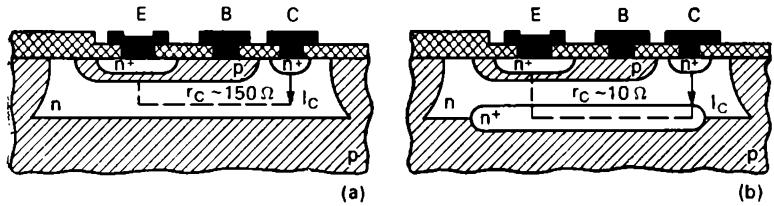


Fig. 5.35. Planar transistor structures without buried layer (*a*) and with buried n^+ layer (*b*)

sive distribution of donors which could cause the n^+ layer and p layer of the base to merge together, the diffusant used for the buried layer must have a low diffusion constant. This can be antimony or arsenic.

At present industry turns out substrates with epi-n layers deposited beforehand, thus relieving the manufacturer of the need to perform the first four operations in the cycle of manufacture of the phototransistor on such a starting substrate.

It is worth noting that the multiple masking, oxidation, and diffusion operations make the profile of the oxide film step-like before metallization. In a number of cases the multistep profile presents difficulties in securing good adhesion of aluminum to the oxide surface. In figures illustrating the structure of a planar transistor, the multistep relief is commonly not shown for simplicity.

The main design distinction of a phototransistor from a common transistor is that the former has a transparent window through which the light gets into the p base as it traverses the thin emitter layer. The need to control the light brings forth the tradeoff between the opposing requirements placed on the structure of a phototransistor. On the one hand, for the phototransistor to be able to effectively generate carriers, the base width w_B must be large. On the other, an increase in w_B is known to reduce the speed and gain of the transistor. As a result, the gain and especially the speed of phototransistors are much lower than those for ordinary transistors. The limiting values of the switching time range from 10^{-6} to 10^{-7} s for the former as against 10^{-9} to 10^{-11} s for the latter; the typical values of β for phototransistors lie between 50 and 200.

5.3.6. Fabrication of photothyristors. The photothyristor has much in common with the ordinary thyristor, the difference being that the former has a window for the incoming light. Fig. 5.36 illustrates the design of a photothyristor in section. The photothyristor can be driven into conduction both by the light penetrating through the window and by the voltage applied to its control electrode to enable it to operate as an ordinary counterpart.

The multilayer *pnpn* structure of a photothyristor can be formed by the diffusion, diffusion-alloying, or planar techniques. As mentio-

ned earlier, the planar process produces all *pn* junctions on one and the same side of the slice. The principal advantage of the planar process is that it forms *pn* junctions under a thin SiO_2 film evaporated over the surface of the slice before the first diffusion step. This film protects the *pn* junctions from contaminants and moisture and serves as a planar passivation layer. Fig. 5.37 illustrates an example of the planar process for manufacturing photothyristors. The basic steps include:

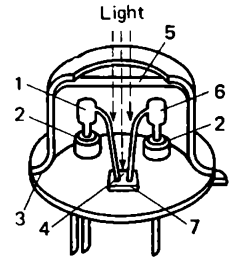


Fig. 5.36. Schematic display of a photothyristor
(*1*) control electrode lead; (*2*) sealed joint; (*3*) weld; (*4*) light-sensitive pad; (*5*) glass window; (*6*) cathode lead; (*7*) silicon chip with *pnpn* structure

(1) deposition of the protective SiO_2 layer (Fig. 5.37*a*);

(2) photomasking and diffusion to form *p* layers for the base and emitter (Fig. 5.37*b*, *c*);

(3) photomasking and diffusion to form the n^+ emitter and preparation of the slice for scribing (Fig. 5.37*d*);

(4) scribing, metallization, and lead interconnection.

Let us point out once more that the photothyristor is a light activated switch whose main purpose is to switch high currents and voltages in power devices such as rectifiers and converters, and in power transmission lines. The planar technique, however, produces photothyristors for use in weak-current devices.

5.3.7. Fabrication of photoresistors. Despite the rapid advances in the domain of *pn* junction photodetectors, photoresistors still hold an important place in optoelectronics. The basic steps in the manufacture of photoresistors are: (1) single crystal growth; (2) powder preparation; (3) vacuum evaporation, or sintering.

Sintering is a widely used technique of thick-film photoresistor fabrication since it is simple and produces films whose parameters are held to close tolerances. The basic steps in sintering, for example, a CdSe-base photoresistor include: preparation of the water-alcohol suspension consisting of a CdSe-CdCl powder blended with an activating $CuCl_2$ additive; suspension spraying on the substrate; air calcination; and deposition of ohmic indium-base contacts.

The technique of vacuum evaporation of thin films can produce photoresistors of a higher speed, with t_{sw} in the range of tens of a microsecond. The vacuum deposition process is well controllable and can use photomasking techniques. A photosensitive material, the CdSe-ZnSe mixture for instance, is evaporated in a vacuum and deposited through a metal mask or a photoresist mask on the substrate. The mask defines the resistor pattern on the circuit. In the next activation stage, the film is calcined in air or argon under the

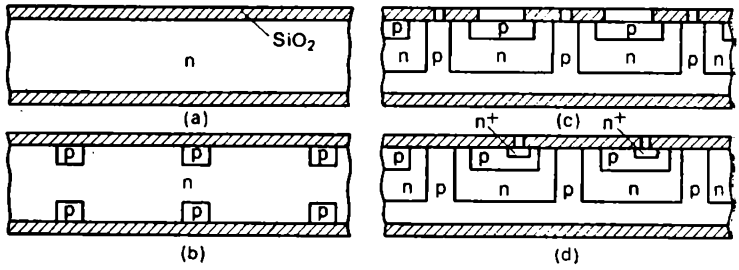


Fig. 5.37. Process of manufacture of a planar photothyristor (*a*) oxidation of starting *n*-type silicon substrate; (*b*) and (*c*) local diffusion to form *p* layers of emitter and base; (*d*) local diffusion to form n^+ layer of emitter

layer of a copper-chlorine powder. The final stage is the deposition of aluminum contacts and bonding pads. The responsivity of evaporated thin-film photoresistors is generally lower than that of sintered thick-film photoresistors. The former also show a poorer reproducibility and lower time stability of parameters. This explains why the sintering technique of photoresistor fabrication still remains very popular.

5.3.8. Fabrication of avalanche photodiodes. These diodes afford high photoamplification and thus present advantages in transferring weak optical signals. Practically all modern diode structures have been tried out for operation in the mode of avalanche multiplication: ordinary *pn* junctions, *pin* diodes, and Schottky barrier junctions. As regards the uniformity of carrier amplification over the entire active area, a specially purified defect-free silicon produces the best results on condition that the diode formed has a guard ring intended to kill the edge effect.

In designing photodiodes, due consideration is given for the fact that holes have a higher avalanche multiplication factor than electrons. That is why the illuminated base region is of the electron conductivity type to enable the generation of hole current. The dark current is largely the current of electrons moving from the *p* region, which have a lower ionizing ability than holes.

Because of a fast increase in the current density during the avalanche breakdown, carrier ionization centers called microplasmas may appear in individual regions of the structure. Microplasmas commonly arise at the edges of the junction due to a growth of the local-field strength and also in the regions of various microdefects both on the surface and in the bulk. Since modern techniques can now produce highly homogeneous structures, microplasmas are most likely to appear at the *pn* junction edges where they come out to the surface, thus inducing the edge effect.

To ensure against the possibility of initiation of the edge effect, a photodiode is made complete with a guard ring—a layer with a lower impurity concentration gradient. As a result, the site of the breakdown shifts from the surface into the bulk. The breakdown region is limited to the central junction region where the strongest field is built up.

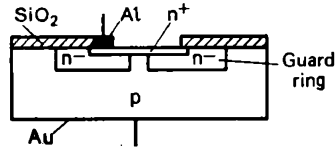


Fig. 5.38. Avalanche photodiode structure

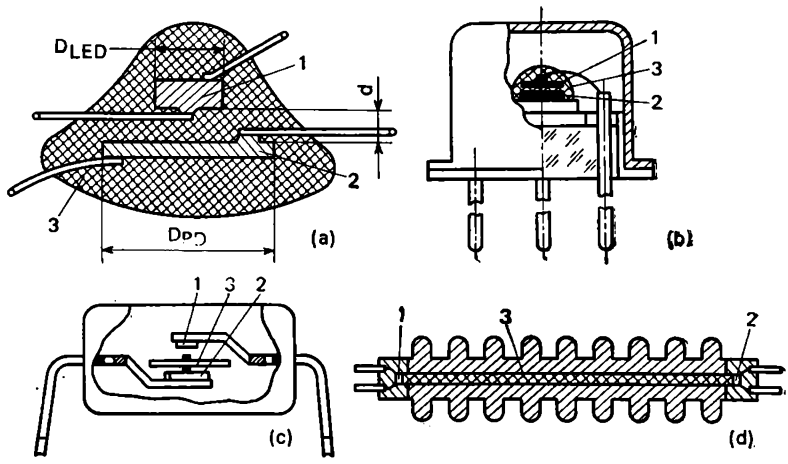
Figure 5.38 displays the structure of an avalanche photodiode with a guard ring. As seen, the structure includes a phosphorus-doped n^+ layer diffused on top of the boron-doped p-type silicon substrate and a guard ring formed by the second diffusion at a lower donor concentration. With a reverse voltage across the diode, the avalanche breakdown develops through the n^+p junction because here the breakdown field builds faster than in the region of the guard ring.

5.4. Fabrication of Photon-Coupled Pairs

5.4.1. The elements of a photon-coupled pair (PCP). As mentioned in Sec. 3.1, the type of photodetector used largely determines the purpose the PCP has to serve. Structurally, however, both the photodetector and photoemitter are equally important. Moreover, it is the emitter that places the highest limits on PCP parameters.

A LED is the basic and most versatile type of light source used in PCPs due to its advantages mentioned in Sec. 1.3. It will be recalled that LEDs are highly efficient electric-to-radiant energy converters capable of operating in a wide spectral range and producing a narrow beam, are fast, small in size, and electrically and technologically compatible with other optoelectronic and electronic devices. The class of photodetectors includes silicon-base photodiodes, phototransistors, photothyristors, etc. The refractive indexes of GaAs, GaAs-base compounds, and silicon are practically the same. So, there is in principle the possibility of securing a perfect match in optical parameters between the photoemitter and PD of the PCP, provided the immersion medium is equal to the job. Consider now the requirements placed on the material of an optical channel (optical medium).

The optical immersion medium must have a high refractive index n, increased resistivity ρ, high critical field strength, good adhesion to silicon and gallium arsenide crystals, adequate elasticity, and improved mechanical strength. Elasticity is a useful property required to match the materials of the photoemitter and photodetector in the coefficient of thermal expansion. The optical channel serves both



ig. 5.39. Photon-coupled pair structures
(a) uncased PCP; *(b)* PCP in metal-glass envelope; *(c)* plastic-encapsulated PCP with lead frame; *(d)* high-voltage optoisolator; *(1)* emitter; *(2)* photodetector; *(3)* optical medium

as a light transmitting medium and as a fastening member of the structure.

The optical medium formed by chalcogenide glass has a high refractive index, $n > 2$. But PCPs using this type of optical path show a low stability to temperature changes (thermal cycles). This is because the chalcogenide-based medium sharply differs from silicon and gallium arsenide in the TC of expansion. Therefore, industrial PCPs mainly utilize organic transparent materials as an immersion medium. These materials do not have refractive indexes in excess of 1.6, but they adhere well to silicon and gallium arsenide and combine a high mechanical strength with enhanced stability to thermal cycles. Wax-like and rubber-like materials find use where optical media need not perform the functions of fastening members.

Thus a PCP consists of a GaAs emitter and silicon PD joined together coaxially by means of the optical medium. Fig. 5.39 illustrates the structures of an uncased PCP (*a*), metal-case packaged PCP (*b*), and plastic-packaged PCPs (*c*, *d*). The structure of Fig. 5.39*d* is shown to have a rather long single-fiber light guide between the photodetector and photoemitter for the purpose of securing a very high input-output isolation voltage $V_{I/O}$, up to tens of kilovolts.

The disadvantage of modern PCPs as regards their manufacturing feasibility is that they are hybrid devices: the PCP consists of discrete elements—a LED and PD fastened together with an optical compound. Such a structure in principle has a low light transfer ability, poor reproducibility of the parameters, small scale of integration,

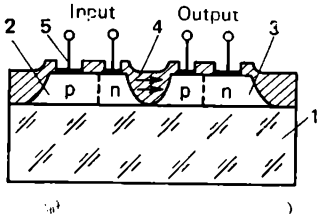


Fig. 5.40. Monolithic PCP structure
(*1*) sapphire substrate; (*2*) emitter; (*3*) detector; (*4*) optical medium; (*5*) leads

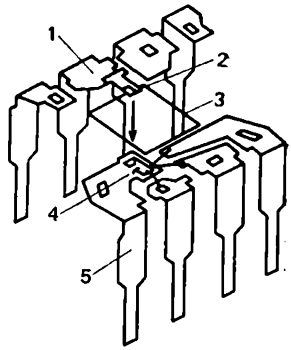


Fig. 5.41. Unencapsulated PCP to be enclosed in a plastic package
(*1*) site for LED mounting; (*2*) LED; (*3*) isolating film; (*4*) silicon photodetector; (*5*) lead

and is not adaptable to batch production, at least, at the stages of PCP mounting and assembling. These shortcomings can only be overcome on converting to monolithic PCPs, one structure version of which appears in Fig. 5.40. Here the light emitter, photodetector, and optical medium are all grown on the same sapphire substrate by the planar-epitaxial method in the single manufacturing cycle. Monolithic PCPs are not yet in commercial production. The ways have still to be found to tackle the difficulties involved in improving the parameters of optical isolation—the purpose the PCP is meant for.

5.4.2. PCP assembling and packaging. Photon-coupled pairs can be encapsulated in molded-plastic packages, ceramic packages, or metal-glass envelopes. Fig. 5.41 illustrates the unencapsulated PCP to be enclosed in a plastic 8-lead dual-in-line pack (DIP). The GaAsP emitter and silicon photodetector are each mounted on an individual plate and connected to four leads. The PCP so assembled with a teflon film sandwiched between the emitter and PD to electrically isolate the input from the output is given a silicone coat. Next the epoxy resin is molded around the pack to seal it and impart rigidity.

In some areas of applications calling for enhanced sealing, molded-plastic packages are not reliable. At present PCPs come in hermetically-sealed metal-glass envelopes or ceramic packages, for example, in a 16-lead metal-ceramic DIP illustrated in Fig. 5.42. The silicon PD is mounted on a ceramic header and connected to its 16-lead frame. The GaAsP emitter with requisite interconnections is fastened on a separate ceramic insert. With solder pellets placed on the header, the insert is pressed onto the heated header to form soldered joints and bond the parts together. To improve optical coupling and electric isolation between the emitter and detector, the pack is coated

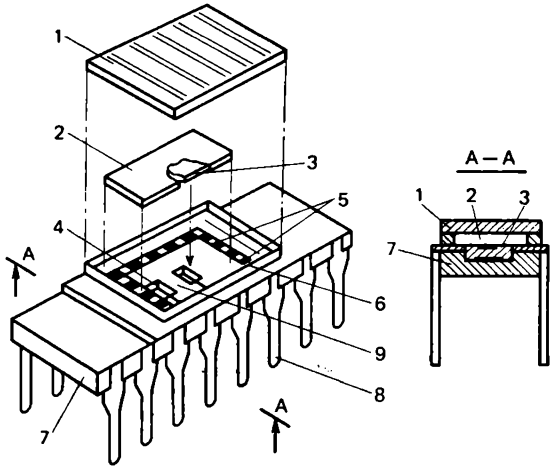


Fig. 5.42. Illustrating the steps of PCP encapsulation in a metal-ceramic envelope
(*1*) metal cap; (*2*) ceramic insert; (*3*) GaAs emitter; (*4*) silicon photodetector; (*5*) solder pellet; (*6*) sealing ring; (*7*) 16-lead ceramic header; (*8*) gold plated leads; (*9*) compound

with a transparent insulation material. The final stage involves sealing with appropriate means. In the tests conducted on such PCP designs, none of the PCPs failed in operation at —55 to +125°C, after storage at —65 to +150°C, and in the environment of 98% relative humidity at +65°C.

5.5. Fabrication of Fiber Guides

5.5.1. Fabrication techniques. In common use are two techniques to produce fiber (light) guides, the method of deposition from the vapor phase and the method of pulling from the dual crucible.

The method of deposition illustrated in Fig. 5.43 relies on the high-temperature reaction to obtain pure quartz SiO_2 from the gaseous mixture of $SiCl_4$ with O_2 at 1200-1600°C:

$$SiCl_4 + O_2 \rightarrow SiO_2 + 2Cl_2$$

The mixture with B_2O_3 or BCl_3 additions is blown through a heated quartz tube to deposit a layer of boron-doped quartz on its wall. Next the core of pure quartz is grown to the required diameter. The tube is then heated to a higher temperature and pressed to clamp the walls. The rod so formed is finally drawn out into a thin filament. The structure of such a rod is shown in Fig. 5.44. Here the core of pure quartz serves as a light conductor; the layer of boron-doped quartz having a somewhat lower refractive index forms a light-refle-

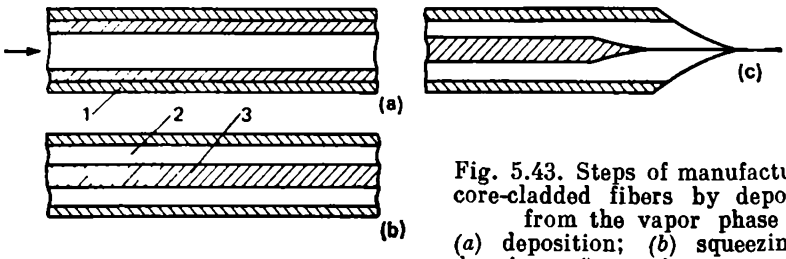


Fig. 5.43. Steps of manufacture of core-cladded fibers by deposition from the vapor phase (*a*) deposition; (*b*) squeezing; (*c*) drawing; (*1*) starting quartz tube; (*2*) shell; (*3*) core

cting shell; the outer layer of the initial quartz tube does not take part in light transmission. The desired light-guide structure will also result if we use the initial quartz tube as a shell and produce a core doped with, say, GeO_2, Al_2O_3, TiO_2, or P_2O_5 which increase its refractive index.

The method of pulling from the dual crucible (Fig. 5.45) serves for extracting a continuous length of fiber from multicomponent glasses whose softening temperature is lower than that of quartz. Glass ingots for forming the light-guide core and the shell are continuously fed in the inner and the outer crucible at the top, while a core-cladded fiber is drawn from the thin orifice at the bottom. To ensure a high purity of the fiber, the arrangement is fitted out with platinum or quartz crucibles, and the extraction zone is made airtight to a certain extent.

One more technique that is suggestive of the above described methods produces the tube and the rod for the core separately; the rod is then inserted into the tube to form the molded block from which the fiber is drawn out. The main disadvantage of this process comes from the difficulty of removing contaminants from the tube and the rod surfaces, which are subsequently brought in contact to form the core-shell interface.

The method of pulling from the dual crucible is also suitable for the fabrication of selfocs; the desired graded change in the refractive index is achieved through the diffusion-induced redistribution of additives in the course of fiber drawing at high temperature. In another widespread method, a homogeneous glass rod is held for a long time in a hot salt melt to obtain the piece to be drawn into a fiber. The ion exchange between the glass and the salt melt decreases the refractive index of the rod's surface layer and thus affords the required gradient of the refractive index.

Compare the two basic methods for fabrication of fiber guides. The method of deposition from the vapor phase enables the highest level of light transmission, but uses only quartz as a starting material and cannot produce a fiber of any desired length. The method of pulling from the dual crucible is free from these limitations—this

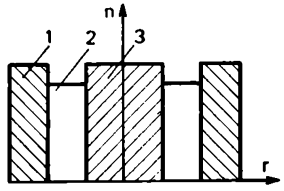


Fig. 5.44. Variation of refractive index n along the radius of the quartz fiber obtained by deposition from the vapor phase (*1*) starting tube; (*2*) shell; (*3*) core

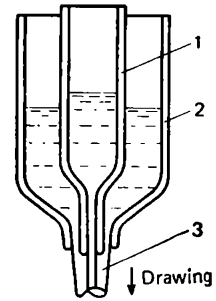


Fig. 5.45. Dual crucible for manufacture of core-cladded fibers
(*1*) inner crucible; (*2*) outer crucible; (*3*) core-cladded fiber

is a continuous process of fiber extraction from any glasses and their combinations. As the technology advances further and the purity of starting glass materials gradually improves, glass fibers become an increasingly closer match to quartz counterparts in transmission parameters.

The quartz fiber fabrication technology has much in common with semiconductor technology—vapor phase deposition that is similar to VPE growth of silicon single crystals, the same range of operating temperatures, and similar material purity requirements (not over 10^{-5}-10^{-6}% impurities).

5.5.2. Fiber guide structures. The choice of the light guide structure is certainly made from considerations of the transmission loss and distortion of an optical signal which is due mainly to dispersion. The equally important factors to be taken into account are the rupture strength of fiber and the resistance to bending and twisting. The rupture strength of quartz and glass fibers is known to be close to 10^{10} Pa. Typical values do not, however, exceed 10^8 Pa. Also, the diameter of bending, D_b, must be not less than 100 to 1000 times as large as the fiber diameter D to exclude mechanical stresses. Proceeding from these conditions, we can conclude that the fiber about 100 μm in diameter is relatively strong and stable to bending and twisting.

It is worthy of note that fibers are not only prone to crack in bending but also to change the properties of light transmission. The transmission loss begins to grow at a radius of bending of merely a few centimeters; increase in optical signal distortion due to enhanced dispersion makes itself felt at D_b of about 1 m.

Figure 5.46 displays the basic structures of a quartz fiber, where n_1 and n_2 are the refractive index of the core and shell respectively.

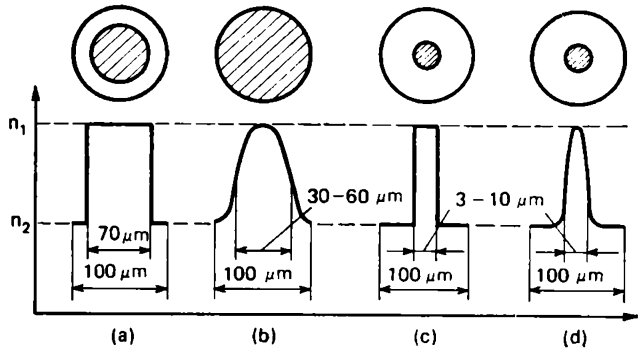


Fig. 5.46. Typical structures of optical fiber
(*a*) multimode core-cladded fiber; (*b*) graded-index fiber; (*c*) single-mode core-cladded fiber; (*d*) graded-index fiber

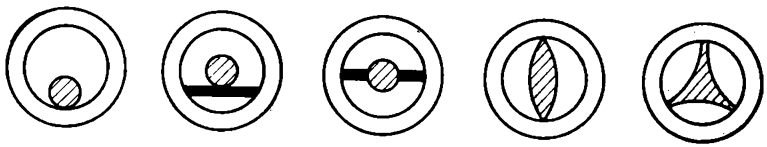


Fig. 5.47. Versions of fiber guides with freely suspended homogeneous quartz cores (shown shaded)

Rather effective for use in light guides is a fiber consisting of a homogeneous quartz core suspended inside a capillary tube. The air inside the capillary plays the role of a light-reflecting shell. Calculations and experiments have defined a variety of optimum core forms which ensure a minimum of light loss and are adaptable to production (Fig. 5.47).

Capillary light guides with a liquid or gaseous filler may prove promising for long-range communication. The advantage of this design is that it can combine a low transmission loss with a large aperture and greater diameter of the light-guide core. An attractive feature from the standpoint of technology is that these light guides can use a wide variety of liquids and gases suitable for light transmission and low-purity glasses. However, the complex structure of such a light guide due to its partial solidity and the instability of its parameters due to infiltration of liquids and especially leakage of gases through thin capillary walls do not yet allow the fluid-filled guide to compete with its quartz counterpart.

Multilayer quartz fibers may prove advantageous for use in long-range communication.

Calculations show that a definite pattern of refractivity changes imparted to individual layers can substantially improve the disper-

sion characteristics of fibers and increase their transmission capacity.

For short-range communication lines, polymer fibers appear to have much promise. In construction and by the principle of action, these fibers are similar to multimode large-aperture glass guides but are much thicker, with D equal to 250-350 μm.

Polymer fiber guides have a lower light transmission capacity; the attenuation factor B is typically about 500 dB/km. The spectral characteristic is satisfactory only up to a wavelength of 0.85-0.90 μm; beyond this range the characteristic displays many absorption peaks caused by CH_3, CH_2; NH_2, and NH_4 groups inherently present in organic compounds as well as by OH ions. The upper limit of the operating temperature range does not exceed 60-80°C. The long-term stability of the parameters is lower than that for glass fiber guides. The attraction of polymer fibers is ease of fabrication combined with low cost as compared to glass fibers.

5.6. Fabrication of Semiconductor Lasers

5.6.1. Laser structure. In Ch. 4 we described the general features, modes of operation, and types of pumping of lasers. Let us recall here that there are a few methods to initiate the lasing action: pumping by way of injection of free carriers into the *pn* junction, electron-beam pumping, and optical pumping.

Carrier injection into the *pn* junction region is the most widely used method of pumping. The advantage of this method comes from its simple implementation combined with high efficiency. As mentioned earlier, the semiconductor laser so pumped is known as the injection laser.

The basic unit of any laser, the injection type including, is its emitter. The simplest lasing emitter is a single laser diode. The *pn* junction of a laser diode must comply with two basic manufacturing requirements, namely, it must be flat and homogeneous to reduce to a minimum the internal loss incurred in modes propagating in the plane of the junction. The end faces must be optically flat and parallel to each other to minimize the edge loss of emission.

Figure 5.48 illustrates the most common *pn* junction with an optical resonator, which satisfies the above requirements. The reflecting faces arranged normal to the junction plane are made optically flat and parallel to each other either by cleaving the crystal along the ⟨110⟩ plane or by polishing. The other two faces are etch treated to banish feedback. Since the refractive index of GaAs is high, the reflectance of faces is fairly high, too, about 0.3 for normal incidence. So there is no need for reflection coats to secure the normal lasing action. To enable a uniform distribution of current, it is of importance to furnish the entire lower and upper faces of the structure

with very good ohmic contacts of low surface resistance; the unevenly distributed current gives rise to the threshold current. Besides, electric contacts must display good thermal properties.

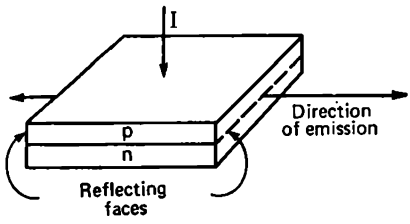


Fig. 5.48. Laser diode structure

Many semiconductor materials are now available for the fabrication of injection lasers. Table 5.3 lists the basic materials and emission wavelengths. In the $Al_xGa_{1-x}As$ semiconductor, the bandgap energy depends on the percentage of components, so the table gives the wavelength range rather than a fixed wavelength.

Injection lasers use semiconductor crystals with *pn* junctions formed by diffusion and epitaxy. Fig. 5.49 illustrates various structures of laser diodes: structure (*a*) formed by diffusion; structure (*b*) formed by epitaxy with one-sided confinement (one-sided heterostructure); structure (*c*) formed by epitaxy with two-sided confinement (double heterostructure).

Table 5.3. **Semiconductor Materials**

Semiconductor	Wavelength, μm	Semiconductor	Wavelength, μm
Gallium arsenide GaAs	0.85	Indium antimonide InSb	5.3
Indium phosphide InP	0.9	Lead telluride PbTe	6.5
Gallium antimonide GaSb	1.6	Lead selenide PbSe	8.5
Indium arsenide InAs	3.2	$Al_xGa_{1-x}As$	0.73-0.85
Lead sulfide PbS	4.2		

5.6.2. Techniques of laser fabrication. Diffusion is the most common technique of forming *pn* junctions for laser diodes. The starting material is *n*-type GaAs doped with tellurium as a donor impurity to a concentration of 10^{18} to 2×10^{18} cm^{-3}. Such an impurity concentration is likely to be optimum for laser diodes with a low level of threshold current. Other donor impurities for GaAs include Se, Si, Ge, and S. Zinc commonly serves as an acceptor impurity. Other acceptors for GaAs are Cd, Mn, and Cu.

The *pn* junction depth reaches 2 to 100 μm depending on the type of starting material, time and temperature of diffusion.

Epitaxial growth methods employed to produce heterojunctions —junctions between two dissimilar semiconductors—now become increasingly popular. The use of heterojunctions enables one to reduce the threshold current density to more than one tenth, produce injection lasers for work in the CW mode at room temperature, and

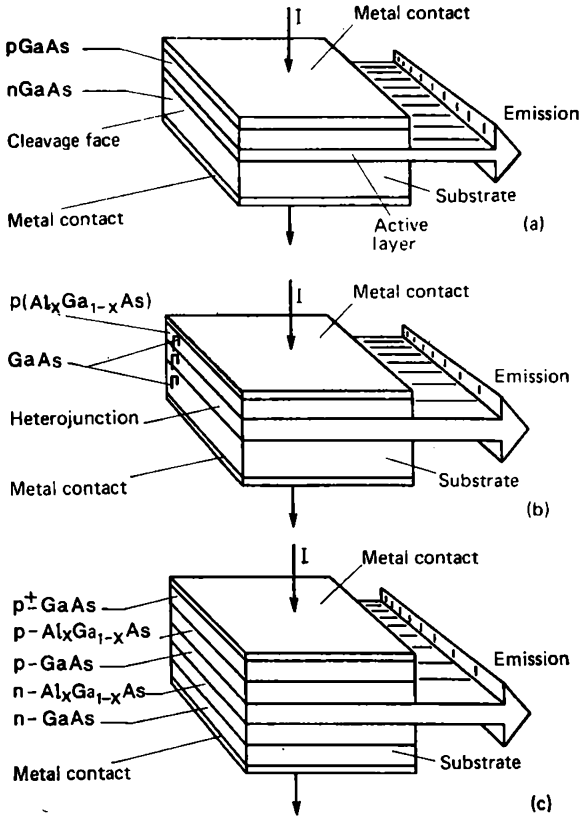


Fig. 5.49. Laser diode structures
(*a*) structure formed by diffusion; (*b*) structure formed by epitaxy with one-sided confinement; (*c*) double heterostructure

sharply increase the life of laser diodes. Heterojunctions originate at the joint between two semiconductors that differ in bandgap energies, but have identical crystal structures and close values of crystal lattice parameters. As a result, the potential barrier appears at the boundary of the heterojunction with sharply improved optical characteristics. This ensures a very close match between the active region generating the light and the region receiving the injected carriers of current. An optimum laser structure is a double heterostructure—a three-layer system with two heterojunctions.

Two semiconductors have proved to be most suitable for the fabrication of heterostructures; namely, GaAs and $Al_xGa_{1-x}As$ which practically have similar lattices but differ in bandgap energies. This pair now goes for the manufacture of GaAs injection lasers with a double heterostructure: n-$Al_xGa_{1-x}As$: p-GaAs: p-$Al_xGa_{1-x}As$.

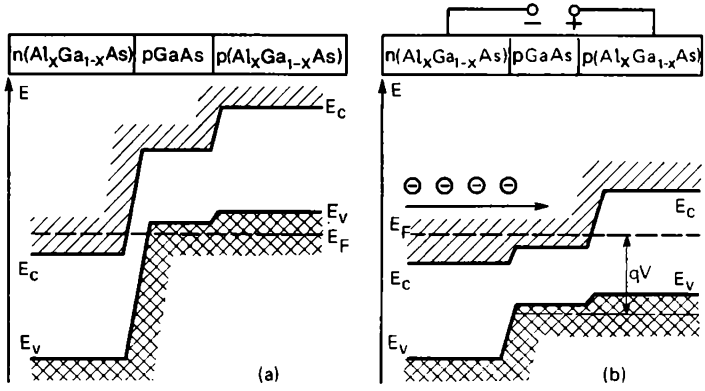


Fig. 5.50. Energy diagrams for a double heterostructure (*a*) in equilibrium state; (*b*) with forward bias

It consists of a thin *p*-GaAs layer about 1 μm thick sandwiched between *p* and *n* regions of $Al_xGa_{1-x}As$ having a larger bandgap.

Figure 5.50 demonstrates a simplified energy diagram of the double structure without a bias (*a*) and with an externally applied forward bias (*b*). The forward bias voltage across the heterostructure causes an injection of electrons from the wide-gap *n*-type $Al_xGa_{1-x}As$ semiconductor to the *p*-type GaAs region, thus providing the population inversion in this region. What accounts for an increased inversion in the active *p*-GaAs layer of the double heterostructure is the so called electron confinement; the electrons injected into *p*-GaAs cannot diffuse into the adjacent wide-gap $Al_xGa_{1-x}As$ region, so their concentration in the active region grows. As seen from Fig. 5.50, the potential barrier at the boundary of the *p*-type $Al_xGa_{1-x}As$ region impedes the diffusion of electrons into this region. Optical properties of heterostructures play an important part in decreasing the threshold current density of heterolasers. The central active *p*-GaAs layer has a higher relative permittivity than extreme *n* and *p* regions of $Al_xGa_{1-x}As$. The active layer with the inversed population thus serves as a dielectric waveguide inside which there propagates a major share of radiation. A lower threshold current density in double heterostructure lasers also stems from the fact that wide-gap $Al_xGa_{1-x}As$ regions are optically transparent to generated radiation.

The best injection GaAs lasers have a threshold current density in the range from 10^3 to 3×10^3 A/cm^2 at room temperature, which is about one hundredth that for the best injection GaAs homolasers. Last years have seen the appearance of heterolasers with an active $Al_xGa_{1-x}As$ layer.

Heterostructures of laser diodes are grown by epitaxial methods.

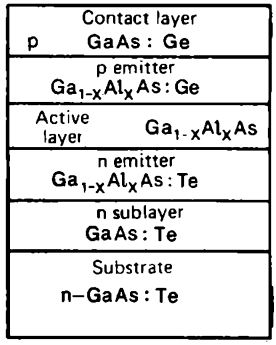


Fig. 5.51. Double heterostructure

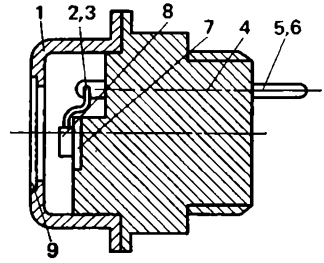


Fig. 5.52. Semiconductor laser structure
(*1*) cap; (*2*), (*3*) copper jumpers; (*4*) base; (*5*), (*6*) leads; (*7*) ceramic plate; (*8*) lasing element; (*9*) transparent window

Small additions of aluminum in the active GaAs region reduce the saturation voltage arising from the mismatch between the lattices of GaAs and adjacent $Al_xGa_{1-x}As$ layers and thus render epi layers more uniform. Aluminum added to the crystal lattice of GaAs increases the gap energy and decreases the emission wavelength. Proceeding from the relation between the optical bandgap and the composition of the solid solution $Al_xGa_{1-x}As$, the required mole concentration of aluminum arsenide in the active region was found; it ranges from 3 to 4% for $\lambda = 855 \pm 5$ nm.

The heterostructures obtained at epitaxial growth temperatures of 870-900°C afford the most reproducible characteristics of laser diodes. This temperature range ensures a relatively low rate of growth and reproducible doping of aluminum in the GaAs lattice. The substrates used for the epitaxial growth of double heterostructures are *n*-type GaAs single-crystal wafers. The structures are grown in an LPE graphite multiple-bin crucible with solution filtration in the atmosphere of pure hydrogen. Since the basic parameters of a laser diode mainly depend on the degree of perfection of the active region, the temperature kinetics of the process calls for the active region growth at a cooling rate of 0.1 K/min. A low rate of cooling promotes the growth of more perfect layers with a low level of defects and also secures a more uniform distribution of aluminum in the active region. The main layers of a heterostructure are grown at a cooling rate of 0.4 K/min.

The thermal treatment of wafers before the epitaxial process causes disturbed layers depleted of arsenic to appear on their surfaces. The disturbed layers are removed by an etchant to a depth of 10 to

20 μm. The schematic representation of an optimum double heterostructure is shown in Fig. 5.51.

The injection laser (Fig. 5.52) is a sealed device which consists of a base *4*, active element *8*, and cap *1*. The laser is assembled in a few basic steps. A beryllium ceramic plate is bonded to the nickel-plated copper base to isolate the lasing element from the base. The lasing element is now mounted on and bonded to the ceramic plate *7*. Current is fed to the lasing element via copper jumpers *2* and *3* which are bonded to the cross-arms of terminals *5* and *6*. The envelope is sealed by cold welding. The cap has a transparent window *9* for light output.

Index

Absorption coefficient, 59, 69
Absorption depth, 59
Applications, optron, 114, 123
  analog, 119, 122
  digital, 114
Area weighted detectivity, 62
Assembling, 172
  photon coupled pair, 207
Azimuth of polarization, 28

Band diagrams, 41, 74
  pin photodiode, 73
  Schottky photodiode, 75
Beam angle, 38, 46, 51
Beam divergence, 25, 135
Birefrigence, 149
Blocking oscillator, 117
Bouguer's law, 59
Brewster angle, 146
Bridgman technique, 177

Capacitance,
  depletion, LED, 51
  input, PCP, 100
  output, PCP, 101
  per-unit area, 87
  transfer, PCP, 104
Cathode-ray tube, 85
CCD, gate,
  operation modes, 88-91
CCD structure, 90, 91
  data recording and readout, 90-92
Character displays,
  parameters, 55
Charge coupled devices (CCDs), 85-92
  operation modes, 88
Charge packet, 88
Chromaticity coordinates, 48
Chromaticity diagram, 49, 50
Coherence, 23, 130
  spatial, 24
  temporal, 24
Color mixture curves, 48
Color purity, 49
Correlation coefficient,
  for optron, 97
Correlation function, 27
Critical wavelength (see also peak wavelength),
  photodetector, 59, 62
Cross-correlation function, 27
Crystal growth techniques, 164-205
Current gain,
  common-base, 81
  common-emitter, 80
Current transfer ratio,
  degradation, 127, 128
  PCP, 102
Cutoff frequency, 23
Czochralski technique, 164, 177

De Broglie equation, 32
Deflectors, 155
  acoustooptic, 156
  electrooptic, 156
Degradation,
  LED, 51, 127
Diffusion constant, 169, 170
Diffusion process, 168, 169, 178
Dominant wavelength,
  LED, 47-50
Doping, 168
Dynistor, see photothyristor

Electric parameters,
  LED, 51
Electroluminescence, 30
  avalanche, 34
  injection, 34, 37
Electromagnetic wave spectrum, 16

Electrooptic effects, 149
Kerr, 150
Pockels, 149, 150
Emission bandwidth, 21, 23
Emission characteristics,
laser, 141
LED, 43, 179
Emission wavelength,
laser, 135, 140
LED, 44
Epitaxy, 165
liquid phase, 166-185
vapor phase, 165-184
Exitance, 18

Fabrication, 164
fiber guide, 208
LED, 174
by diffusion, 178-181
by liquid phase epitaxy, 185, 191
by vapor phase epitaxy, 182, 191
ohmic contact, 188
photodetector, 196-205
materials for, 196
photon coupled pair, 205
semiconductor laser, 212
materials for, 213
Fabrication processes, 164-205
Fermi-Dirac function, 139
Fiber guides, 159
fabrication, 208
numerical aperture, 159
structures, 210
transmission loss in, 160
Fiber-optic communication lines, 157
Fick's laws, 169
Fluorescence, 30

Heterodiodes, 197
fabrication, 199
Heterojunctions, 40-43
band diagrams, 41
diode, 76, 197-199
double, 41, 43
single, 41, 43
spectral characteristics, 42
Heterolasers, 215
Heterostructures (see also heterojunctions), 215, 216
Holographic devices, 163
Holographic memory, 162

Illuminance, 16, 19
Injection ratio, 36, 38
Ion implantation, 168, 169
Irradiance, 16, 19
Isolation resistance,
PCP, 104
Isolation voltage,
PCP, 103

Lasers, 129
continuous wave, 137
gas, 144
arc-discharge-excited, 145
glow-discharge-excited, 146
modes of operation, 137
parameters, 137, 147
pulsed, 137
semiconductor, 138
electron-beam, 138, 143
fabrication, 212
heterojunction, 143
injection, 138-143
solid-state, 146
Laser modulation, 136, 149
Laser modulators, 148
Laser pumping, 130
chemical, 134
electric-field, 134
electron-beam, 134, 143
longitudinal, 144
transverse, 144
optical, 134
Lasing, 133
threshold, 134
Law of conservation of momentum, 32
Light-activated SCS, see photothyristor
Light amplification, 130, 131
Light emitting diodes (LEDs), 15-56
beam angle, 46
degradation, 51, 127
depletion capacitance, 51
diffused, 53, 181
dominant wavelength, 46
emission characteristics, 43
emission wavelength, 46, 52
epitaxial, 53
fabrication, 174
heterojunction, 40
injection, 95
materials for, 192
mesa, 193
planar, 193
quantum efficiency, 38
radiation loss in, 39, 40
radiation pattern, 38, 51, 54
radiation spectrum, 45
spectral characteristics, 44

switching speed, 51
threshold current, 44
VI characteristics, 52
Light guide, see fiber guide
Liquid phase epitaxy, 166
apparatus for, 185-188
Long-wavelength cutoff, 59
photodetector, 69
Schottky diode, 76
Luminance, 16, 19
Luminescence, 30
semiconductor, 29-33
Luminosity, 17
Luminous efficacy, 17
Luminous efficiency,
of the eye, 17, 45, 48
Luminous exitance, 16
Luminous flux, 16
Luminous intensity, 16
LED, 52

Magnetooptic effect, 150
Mesa-structure, 42, 193, 195
Modulation factor, 152
Modulators, optical, 151
electrooptic, 151
magnetooptic, 153, 154
semiconductor, 153
space-time, 156
thin-film, 153
Momentum,
electron, 32
photon, 32
Monochromaticity, 21, 23, 24
MOS capacitor, 86
Mutual coherence function, 25-27

Noise, photodetector, 61
Noise equivalent power (NEP)
specific, 62
unit bandwidth, 62

Optical cables,
designs, 161
Optical parameters,
LED, 46
Optical output efficiency, 97
LED, 38, 127
Optical radiation, 15
coherence, 23
emissive power, 29
loss, 39, 160
luminescent, 29
photometric parameters, 15, 20
polarization, 28
polarized, 28
pulse power, 29
radiometric parameters, 15, 20
self-absorption, 33
spectral analysis, 23
Optical resonator, 133
photon avalanche in, 134
Optical signal attenuation, 160
Optoelectronic devices,
fabrication, 164-214
Optrons, see also PCPs, 42, 93
analog, 114
application, 114
block diagram, 94
classification, 99
differential, 121
digital, 114
fabrication, 205
fiber-optic channel, 109
optical medium for, 96, 108
parameters, 100-109
photodetectors for, 98
photoemitters for, 95
properties, 93
special, 99, 108, 123
Optron switches, 123
Optronic ICs, 99, 107
analog, 119, 121
digital, 119
Oxidation, 167

Packaging,
PCP, 207
Peak wavelength,
LED, 47, 52, 192
Penetration depth, 59, 69
Phosphorescence, 30
Photochromatic effect, 151
Photoconductive effect, 57, 83, 151
Photocoupled isolators, see PCPs
Photocurrent, 63-71
gain, 77-81
transfer ratio, 81
Photodarlington, 79, 80, 98
Photodetectors, 45, 56-59
CCD-based, 85
characteristics, 62
diode, 63-78, 196-204
fabrication, 196-204
for optrons, parameters, 98, 101
noise in, 61
parameters, 59, 61
resistor, 81, 82
responsivity, 60, 61
spectral response, 60, 63

transistor, 79-98
Photodiodes, 63
avalanche, 77, 204
VI characteristics, 78
fabrication, 196-204
heterojunction, 76
modes of operation,
photoconductive, 66, 71
photovoltaic, 66, 71
photoeffect in, 64
pin, 72, 196
Schottky barrier, 74, 198
transients in, 71
VI characteristics
Photoeffect parameters, 59
Photoelectric effect,
external, 57
internal, 56, 64
photoconductive, 57, 58
photovoltaic, 57, 58
Photoelectromotive force, 57, 65, 71
Dember, 65
Photoemissive effect, 57
Photoemitters, see also lasers and LEDs,
semiconductor,
characteristics, 192
materials for, 192
structures, 195
Photolithography, 170, 200, 201
Photoluminescence, 30
Photometric parameters, 15, 20
definition, 15-20
Photon coupled pairs (PCPs), 99
commercial, 109
parameters, 110-112
current transfer ratio, 102
diode, 104, 110
fabrication, 205
packaging, 207
parameters, 100-107, 120
quality factor, 107
resistor, 107, 112
speed of response, 103
thyristor, 107-113
transfer characteristics, 102, 105, 121
transients in, 102
transistor, 106, 111
Photoresists, 170
Photoresistors, 81, 82
fabrication, 203
Photothyristors, 81, 98
fabrication, 202-204
Phototransistors, 79, 98
current gain, 80
fabrication, 200
output characteristics, 80
responsivity, 80
Planck law, 29
Pockels effects, 149, 150
Polarization,
elliptical, 28
linear, 28
Population inversion, 139
definition, 132
Pulse shaper, 124

Quantum efficiency, 59
external, 38, 141
internal, 37, 38, 68
laser, 141
LED, 38, 192

Radiance, 16, 18
Radiant exitance, 16
spectral, 29
Radiant flux, 16, 17
Radiant intensity, 16
Radiation, see optical radiation
Radiation loss, 160
backscatter, 40
edge, 40
Fresnel, 39, 97
self-absorption, 39
TIR, 39
Radiation pattern, 38
laser, 135, 141
LED, 51, 54
Radiation spectra, 23
Radiation stability, 128
Radiators (see also LEDs), 15, 38
semiconductor, materials for, 33
Radiometric parameters, 15, 20
definitions, 15-19
Ramp generator, 118
Random function, 26
Recombination, 32, 139
across *pn* junction, 36
Auger, 37
radiative, 33
through luminescent centers, 36
Reliability, parameters, 125
Reliability criteria, 126, 127
Reliability function, 125
Responsivity,
photodetector, 60, 61, 73

Selfoc, 159, 160
fabrication, 209
Semiconductors,

direct-gap, 31, 32
indirect-gap, 31, 33
light absorption in, 59
narrow-gap, 40
photoluminescence in, 31
wide-gap, 40, 174
Shot-wavelength cutoff,
photodetector, 63, 69
Schottky diode, 76
Single crystal growth, 175
Solid angle, 16, 17
Solid solubility, 168
Spatial coherence function, 24-26
Spectral characteristics,
filter, 48
hot body, 30
LED, 44
photodetector, 60
photodiode, 69
Spectral concentration, 22
of power, 29
Spectral luminous efficiency,
of the eye, 17, 45, 48
Spectral parameters, 21
Spectral response,
photodetector, 63, 69
Spectrum locus, 49
Speed of response,
LED, 51, 52
PCP, 103
photodiode, 70
Switching time,
LED, 51
PCP, 106, 109
photodetector, 63
Temporal coherence function, 24
Thermocompression bonding, 173
Threshold current,
laser, 142
LED, 44
Time lag, LED, 51
Total internal reflection (TIR), 39, 159
Transfer characteristics, PCP, 102, 105
Transitions, 131, 132
direct, 32, 33
indirect, 32, 33
interband, 33
Transmission loss, 160
Transmittance, filter, 47
Tristimulus values, 48

Vapor phase epitaxy, 166, 167, 182
apparatus for, 182-184
Voltage-current characteristics
forward, 51
LED, 51, 52
photodiode, 67, 78
photoresistor, 81
photothyristor, 82
reverse, 51

Waveforms, 90, 102, 117, 118
photocurrent, 71

TO THE READER

Mir Publishers welcome your comments on the contents, translation, and design of the book.

We would also be pleased to receive any suggestions you care to make about our future publications.

Our address is:
USSR, 129820,
Moscow, I-110, GSP,
Pervy Rizhsky Pereulok, 2,
Mir Publishers

For your library

V. V. Meshkov

FUNDAMENTALS OF ILLUMINATION ENGINEERING

There are numerous problems which arise in connection with the methods of design and calculation of light sources and lighting installations and these are dealt with in detail here. This book is a comprehensive treatise on illumination engineering. The many important topics are discussed in three chapters: Radiometric and Effective Characteristics of Optical Radiation; Elements of Radiometry; and Conversion of Radiant Energy. There are several sections in each chapter that cover practical lighting techniques and calculation.

The audience for the book will include college students and research workers in illumination engineering.

The book was published by MIR Publishers in 1981.